

Modellierung der Bindungsregionen von Protein-Protein-Komplexen unter besonderer Berücksichtigung von Wassermolekülen

Vom Fachbereich Chemie
der Technischen Universität Darmstadt

zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigte
D i s s e r t a t i o n

eingereicht von

Dipl.-Ing. Thorsten Borosch
aus Groß-Umstadt

Berichterstatter:	Prof. Dr. J. Brickmann
Mitberichterstatter:	Prof. Dr. R. Schäfer
Tag der Einreichung:	15. Februar 2005
Tag der mündlichen Prüfung:	11. April 2005

Darmstadt 2005

*Für Oma und im
Andenken an meine Familie.*

Die vorliegende Arbeit wurde im Fachbereich Chemie der Technischen Universität Darmstadt, Fachgebiet Physikalische Chemie I, unter der Leitung von Prof. Dr. J. Brickmann in der Zeit von April 1999 bis Januar 2005 durchgeführt.

An dieser Stelle danke ich besonders:

Prof. Dr. Jürgen Brickmann für seine Unterstützung und die mir gewährte Freiheit bei der Durchführung der Arbeit;

Dr. Hans-Jürgen Bär für die Durchsicht der Arbeit und seine unermüdliche Hilfe bei allen organisatorischen Fragen;

Dr. Thomas Exner für die vielen hilfreichen Diskussionen und Anregungen bei der Durchsicht der Arbeit;

Dr. Matthias Keil für die Bereitstellung und Einführung in viele seiner Programme, für die sonntägliche Motivation sowie für seine allzeit hilfsbereite Art;

Dr. Stefan Kast und Dr. Marco Müller für die wissenschaftlichen Diskussionen und so manch neue Idee für die Arbeit;

Dr. Peter Monecke und Dipl.-Ing. Bernd Schilling für die Bereitstellung ihrer Wasserdaten sowie Dipl.-Inf. Oliver Korb und Jens Gimmler für ihre Programme;

Dipl.-Chem. Thorsten Schepers für seine freundschaftliche Art und die Musik;

allen Mitgliedern des Arbeitskreises Physikalische Chemie I für die sehr gute Arbeitsatmosphäre und den „Plausch am Rande“;

der Deutschen Forschungsgemeinschaft für die Förderung von Teilen dieser Arbeit;

sowie Susanne Heger für die stete Unterstützung, Motivation und ihre Bereitschaft, immer ein offenes Ohr für mich zu haben.

Inhaltsverzeichnis

1 Einleitung	1
2 Grundlagen	4
2.1 Physikalische und strukturelle Eigenschaften von Proteinkomplexen	4
2.2 Wassermoleküle in der Binderegion von Proteinkomplexen.....	8
2.3 Vorhersage der Affinität von Proteinkomplexen	11
2.4 Molekulare Oberflächen	13
2.4.1 Die van-der-Waals-Oberfläche	14
2.4.2 Lösungsmittelzugängliche Oberfläche.....	14
2.4.3 <i>Separating Surface</i>	16
2.5 Darstellung physikochemischer Daten auf molekularen Oberflächen.....	17
2.5.1 Elektrostatisches Potential	18
2.5.2 Lokale Lipophilie.....	20
2.5.3 Wasserstoffakzeptoren und -donatoren.....	23
2.5.4 Salzbrücken.....	25
2.5.5. Topographische Analyse der Oberfläche	26
3 Vorhersage von Wasserpositionen in Proteinkomplexen	29
3.1 Auswahl des Datensatzes.....	29
3.2 Bearbeitung der Kristallstrukturen.....	32
3.3 Beschreibung des Verfahrens zur Vorhersage von Wasserpositionen.....	33
3.4 Parametrisierung des Verfahrens	36
3.4.1 Erstellung eines Datensatzes an Kavitäten mit optimaler Wasseranzahl.....	37
3.4.2 Optimierung des Radius der „Wasserkugel“	40
3.4.3 Einfluß des „äußeren“ Radius.....	43
3.5 Optimierung des Auffüllvorganges mit der Erweiterung <i>sphopt</i>	46
3.6 Vergleich des Verfahrens mit dem Programm <i>DOWSER</i>	48
3.7 Validierung des Verfahrens mit Kristallwassermolekülen	50
3.8 Diskussion der Ergebnisse des <i>space</i> -Verfahrens.....	52
4 Statistische Analyse der Binderegion	54
4.1 Berechnung physikochemischer und geometrischer Daten	54
4.1.1 Bestimmung und Unterteilung der Oberflächenanteile der Binderegion.....	55
4.1.2 Berechnung physikochemischer/geometrischer Daten auf der Proteinoberfläche	56

4.1.3 Grundlage der statistischen Auswertung	58
4.1.4 Berechnung von Volumen und Flächeninhalt.....	58
4.2 Analyse der Kavitäten- und Wasserverteilung in der Binderegion.....	59
4.3 Analyse der physikochemischen Eigenschaften	66
4.3.1 Elektrostatisches Potential (ESP).....	66
4.3.2 Lokale Lipophilie (MLP).....	68
4.3.3 Wasserstoffbrücken-Donoren/Akzeptoren (HB2) und ihre Dichte (HD)	68
4.3.4 Salzbrückenbildner (SB2).....	70
4.3.5 Tiefeninformation (CAV)	70
4.3.6 <i>Shapeindex</i> (SHI) und <i>Curvedness</i> (CUR)	72
4.4 Zusammenfassung der statistischen Analyseergebnisse	74
5 Vorhersage von Bindungskonstanten	77
5.1 Auswahl des Datensatzes	78
5.2 Berechnung der Deskriptoren	80
5.2.1 Berechnung der <i>Separating Surface</i>	81
5.2.2 Elektrostatisches Potential D_{esp}	82
5.2.3 Lokale Lipophilie D_{mlp}	82
5.2.4 Salzbrücken D_{sb2} und Wasserstoffbrücken $D_{\text{hb2}} / D_{\text{nh}}$	83
5.2.5 Kavitäten- und Wasserinformation $D_{\text{nk}}, D_{\text{nw}}, D_{\text{k}}$	84
5.2.6 Größe unterschiedlicher Binderegionsbereiche $D_{\text{a}}, D_{\text{PPa}}, D_{\text{al}}, D_{\text{PPal}}$	85
5.3 Auswahl optimaler Deskriptoren	85
5.4 Untersuchung des aufgestellten Modells	88
5.5 Diskussion des aufgestellten Modells.....	92
6 Zusammenfassung und Ausblick	94
7 Literatur	98
8 Anhang	109
8.1 Ergänzende Tabellen zur statistischen Auswertung	109
8.1.1 Größe der Binderegion, Wasser- und Kavitätenanzahl des Datensatzes	109
8.1.2 Volumen, Oberfläche und Wasserbesetzung der Kavitäten des Datensatzes	112
8.2 Ergänzende Tabellen zur <i>Scoring</i> -Funktion	120
8.2.1 Deskriptorenwerte der <i>Scoring</i> -Funktion	120
8.2.2 Experimentelle und berechnete Bindungsaffinitäten	121
8.3 Verwendete Hilfsmittel.....	122

1 Einleitung

Die zentrale Bedeutung von Proteinen für Prozeßabläufe innerhalb der Zelle wurde bereits in der Anfangszeit biochemischer Forschung erkannt. Proteine sind in der Lage, sehr unterschiedliche Aufgaben innerhalb eines Organismus zu übernehmen [1]. Faserproteine wie Kollagen in Knochen und Sehnen oder Keratin in Haut und Haaren dienen als Strukturmatrix lebender Organismen und sind passive Strukturbildner. Andere Proteine wie Myosin und Actin sind bei der Bildung von Muskelfasern beteiligt und können koordinierte mechanische Bewegungen erzeugen. Proteine dienen als Regulatoren zur Aufrechterhaltung der gesamten Zellfunktion: in Form chemischer Botenstoffe zur Signalübertragung (Hormone), als Rezeptoren dieser Botenstoffe, als Transport- und Speicherproteine biologisch relevanter Substanzen, als Katalysatoren chemischer Reaktionen bei Stoffwechselvorgängen (Enzyme) oder zur Abwehr von Infektionserregern als zentraler Bestandteil des Immunsystems höherer Organismen (Immunglobuline).

Die Funktionalität der Proteine ist eng mit der Eigenschaft verknüpft, nicht-kovalente Bindungen mit anderen Molekülen eingehen zu können und Molekülkomplexe zu bilden. Spezifität und Affinität der Komplexbildung sind für die Regulation physiologischer Prozesse entscheidende Faktoren und werden sehr stark durch die dreidimensionale Struktur der beteiligten Komplexpartner bestimmt. Die strukturelle Komplementarität der Komplexpartner wurde von Emil Fischer bildhaft als Schlüssel-Schloß-Prinzip beschrieben [2]. Struktur, Stabilität und Funktionsweise von Proteinen werden maßgeblich durch das umgebende Medium Wasser mitbestimmt. Das Wasser ist nicht nur Lösungsmittel, sondern stellt in vielen Fällen einen integralen Bestandteil der Komplexbinderegion dar, und trägt durch vernetzende Wasserstoffbrücken zur strukturellen und physikochemischen Komplementarität der Komplexpartner bei.

Die dreidimensionalen Strukturen mehrerer tausend Proteine und Proteinkomplexe mit nicht-kovalent gebundenen Wassermolekülen konnten in den letzten Jahren mit Hilfe von Röntgenstrukturanalysen und NMR-Experimenten bestimmt werden. Die Strukturdaten sind in der *Protein Data Bank* (PDB) [3] gespeichert und frei zugänglich. Sie dienen als Basis zum Verständnis der strukturellen und physikochemischen Eigenschaften, die eine Komplexbildung ermöglichen. Zur Bestimmung dieser Eigenschaften werden neben experimentellen auch rechnerbasierte Methoden verwendet. Diese reichen von der

Visualisierung der hochkomplexen dreidimensionalen Proteinstrukturen bis zur Simulation ganzer Proteinkomplexe in ihrem Lösungsmittel [4]. Die Aufklärung von Protein-Protein-Interaktionen ist ein wesentliches Teilgebiet der *Proteomik*, welche die Erforschung der Stoffwechsel- und Regulationswege aller in einer Zelle exprimierten Proteine zum Ziel hat.

Besonders bei der Pharmaforschung ist ein grundlegendes Verständnis der die Spezifität und Affinität bestimmenden Faktoren von Proteinkomplexen für die Entwicklung neuer Medikamente von Bedeutung [5]. Hier wird nach Strukturen gesucht, die eine hohe Bindeaffinität zu einem bestimmten Enzym haben, dessen physiologische Funktion z.B. gehemmt werden soll. Pharmaka können aber auch als *Agonisten* wirken, indem sie durch Bindung an einen Rezeptor den physiologischen Liganden imitieren und so eine Rezeptorantwort auslösen. Ein experimenteller Ansatz zur Identifikation von Wirkstoffen ist das *High Throughput Screening*. In einem automatisierten Verfahren wird für eine Vielzahl von niedermolekularen Substanzen und Naturstoffen (sog. Substanzbibliotheken) die Bindeaffinität zum Zielprotein bestimmt. Bildet das Zielprotein in seiner physiologischen Umgebung Protein-Protein-Komplexe aus, kommen kleine, nicht-peptidische Liganden, wie sie in den meisten Substanzbibliotheken enthalten sind, als Wirkstoffe meist nicht in Frage, da ihre Bindeaffinität zu gering ist. Die Untersuchung von Protein-Protein-Wechselwirkungen zur Entwicklung geeigneter Peptide als Wirkstoffe gewinnt daher immer stärker an Bedeutung [6]. Um den experimentellen Aufwand bei der Suche nach Wirkstoffen möglichst gering zu halten, werden im Vorfeld rechnergestützte Methoden verwendet, die es erlauben, anhand struktureller und physikochemischer Eigenschaften der Komplexpartner die Bindeaffinität des Komplexes vorherzusagen. Die Entwicklung guter Computermodelle, mit denen Wirkstoffe *in silico* entworfen oder aus einem großen Pool bekannter Substanzen herausgefiltert werden können, erfordert daher ein breites Verständnis der die Komplexbildung bestimmenden Faktoren.

An der Komplexbildung sind neben den Proteinen meist auch Wassermoleküle beteiligt, da fast alle physiologischen Reaktionen in wässrigem Medium ablaufen. So konnten in der Binderegion von Protein-Protein-Komplexen Wassermoleküle nachgewiesen werden, die einen erheblichen Beitrag zur Bindeaffinität und zur strukturellen Komplementarität der Komplexpartner leisten und destabilisierende Hohlräume füllen [7,8]. Da diese Wassermoleküle in Kavitäten der Binderegion noch immer ein hohes Maß an Mobilität besitzen, können sie bei Röntgenstrukturanalysen nur schwer oder gar nicht nachgewiesen

werden [9]. Aufgabe der vorliegenden Arbeit war es, ein Verfahren zu entwickeln, mit dem die Anzahl und die potentiellen Positionen solcher Wassermoleküle in Hohlräumen der Binderegion von Proteinkomplexen automatisch vorhergesagt werden können. Dazu mußte ein neuer Algorithmus entwickelt werden, der auf einem geometrischen Ansatz beruht. Der Algorithmus wird in Kapitel 3 ausführlich beschrieben.

Die Vorhersage potentieller Wasserpositionen in der Binderegion von Komplexen sollte die Möglichkeit bieten, eine statistische Analyse dieser Wassermoleküle an einem großen Datensatz von Proteinkomplexen bekannter Kristallstruktur durchzuführen, auch wenn keine Informationen über Kristallwassermoleküle vorhanden sind. Das Ziel der in dieser Arbeit durchgeführten statistischen Analyse sollte es sein, Unterschiede zwischen Teilbereichen der Binderegion mit Protein-Wasser- und Protein-Protein-Kontaktflächen hinsichtlich lokaler struktureller und physikochemischer Kenngrößen aufzudecken. Zum Vergleich der unterschiedlichen Kontaktbereiche sollte das Konzept molekularer Oberflächen verwendet werden. Es ermöglicht die lokale Trennung der Kontaktbereiche sowie die Verbildlichung von lokalen Moleküleigenschaften auf diesen Oberflächen. Vorgehensweise und Ergebnisse der Analyse werden in Kapitel 4 besprochen.

Die Kenntnis über die potentielle Anzahl und Verteilung von Wassermolekülen in der Binderegion kann nicht nur zum besseren Verständnis von Protein-Protein-Wechselwirkungen beitragen, sondern auch zur Verbesserung bestehender Methoden zur Vorhersage von Bindeaffinitäten genutzt werden. Eine schnelle und daher weitverbreitete Methode zur Vorhersage von Bindeaffinitäten ist die Verwendung empirisch optimierter Funktionen, mit denen die Bindungskonstante meist durch lineare Kombination von strukturellen und physikochemischen Eigenschaften der Komplexpartner bestimmt werden kann (sog. *Scoring*-Funktionen) [10]. In dieser Arbeit sollte erstmals eine *Scoring*-Funktion für Protein-Protein-Komplexe entwickelt werden, in der die Information über die vorhergesagten Wassermoleküle bzw. Hohlräume in der Binderegion berücksichtigt wird. In Kapitel 5 wird dargestellt, wie die Information der Wassermoleküle in die *Scoring*-Funktion integriert werden kann und welche weiteren Deskriptoren sich als sinnvoll erwiesen haben. Eine kurze Zusammenfassung aller Ergebnisse und ein Ausblick auf mögliche Weiterentwicklungen wird schließlich in Kapitel 6 gegeben.

2 Grundlagen

2.1 Physikalische und strukturelle Eigenschaften von Proteinkomplexen

Eine Vielzahl biologischer Funktionen in der Zelle hängt von der Fähigkeit von Proteinen ab, mit anderen Proteinen Komplexe auszubilden. Protein-Protein-Wechselwirkungen sind sowohl an strukturellen Aufgaben wie der Instandhaltung des Cytoskelettes, als auch an funktionellen Prozessen wie der Übermittlung zellulärer Signale (Signaltransduktion) oder der Kontrolle der Genexpression beteiligt. Fehlfunktionen dieser Wechselwirkungen sind häufig die Ursache unterschiedlichster Krankheitsbilder. Es gibt daher Ansätze zur Entwicklung von Molekülen, die Protein-Protein-Wechselwirkungen inhibieren (sog. *Antagonisten*) und für therapeutischer Zwecke angewendet werden könnten [6,11]. Zur erfolgreichen Entwicklung solcher Antagonisten ist es notwendig, die Ursachen von Spezifität und Affinität bei Protein-Protein-Wechselwirkungen zu kennen. Trotz zahlreicher Untersuchungen sind die Prinzipien der Protein-Protein-Erkennung immer noch nicht vollständig verstanden.

Startpunkt fast aller Untersuchungen sind die experimentell bestimmten, dreidimensionalen Kristallstrukturdaten von Proteinkomplexen, die in der *Protein Data Bank* [3] frei zugänglich sind und deren Anzahl in den letzten Jahren stark zugenommen hat. Basierend auf diesen Daten lassen sich Proteinkomplexe in unterschiedliche Komplextypen einteilen [12]. Proteine, die nur als Oligomer existieren und nur als solches funktionell wirken (meist Homodimere) werden im folgenden als *obligatorische* Komplexe bezeichnet. Davon zu unterscheiden sind Komplexe aus Proteinen, die sowohl als stabiles Monomer als auch komplexiert als Oligomer vorkommen können (z.B. Antigen-Antikörper, Enzym-Inhibitor). Jones *et al.* konnten zeigen, daß sich die unterschiedlichen Typen in ihren Bindungseigenschaften unterscheiden lassen [13,14]. Darüber hinaus gibt es auch Studien über nicht-spezifische Protein-Protein-Wechselwirkungen, wie sie etwa in Kristallpackungen vorkommen [15,16]. Diese Art von Proteinkomplexen wird in der vorliegenden Arbeit nicht untersucht, da die intermolekularen Wechselwirkungen durch die Kristallpackung „aufgezwungen“ werden und daher keine große Übereinstimmung mit dem physiologischen Protein-Protein-Erkennungsprozeß aufweisen.

Zum besseren Verständnis von Protein-Protein-Wechselwirkungen wurden von unterschiedlichen Autoren Untersuchungen der strukturellen und physikalischen Faktoren durchgeführt, die wesentlich an den Protein-Protein-Wechselwirkungen beteiligt sind und diese beeinflussen:

- Form, Größe und Aminosäurezusammensetzung der Binderegion [13,17-22]
- Hydrophobe Wechselwirkungen in der Binderegion [13,17,18,20,23-25]
- Elektrostatische Komplementarität, Salz- und Wasserstoffbrücken [13,17,18,26-30]
- Entropieänderungen bei der Komplexbildung [31,32]

Voraussetzung zur Bildung eines stabilen Protein-Protein-Komplexes ist die optimale Paßform der Komplexpartner. Jones *et al.* [13] konnten mit dem von ihnen definierten *Gap*-Index, dem Volumen von Hohlräumen zwischen den Komplexpartnern bezogen auf die Bindungsoberfläche, eine hohe sterische Komplementarität bei Homodimeren und Enzym-Inhibitor-Komplexen nachweisen. Im Vergleich dazu ist die sterische Komplementarität von Antigen-Antikörper-Komplexen weniger stark ausgeprägt [21], was auf die unterschiedliche evolutionäre Entwicklung und chemische Natur der Bindestelle von Antikörpern zurückgeführt werden kann. Die Größe der mittleren Kontaktfläche von Proteinkomplexen liegt bei ca. $1600 (\pm 400) \text{ \AA}^2$ für den Gesamtkomplex [17]. Die Binderegion eines Monomers entspricht also etwa 10% der Gesamtoberfläche des Monomers. An den intermolekularen Kontakten sind im Durchschnitt 170 Atome pro Komplexpartner beteiligt, die sowohl zum Proteinrückgrat als auch zu den Seitenketten gehören können. Die Zahlen sind nur als grobe Näherung zu verstehen und sollen einen Eindruck über die vorliegenden Größenordnungen vermitteln. Die tatsächliche Größe der Binderegionen ist sehr heterogen und hängt stark vom Typ des betrachteten Komplexes ab. In dem von Conte *et al.* [17] untersuchten Datensatz reichte die Größe der Binderegionen von 1150 \AA^2 bei kurzlebigen, schwachen Komplexen (z.B. Cytochromperoxidase-Cytochrom c) bis zu 4660 \AA^2 bei Komplexen, die am Signaltransport (G-Proteine) beteiligt sind. Konformationsänderungen der Komplexpartner während der Komplexbildung sind meist bei Komplexen mit großen Kontaktbereichen ($2000\text{-}4000 \text{ \AA}^2$) zu beobachten [17]. Binderegionen der „Standardgröße“ weisen nur geringe Änderungen der Proteinkonformation auf.

Neben der geometrischen Paßform der Komplexpartner trägt auch die Aminosäurezusammensetzung in der Binderegion erheblich zur Komplexstabilität bei. Untersuchungen der Aminosäureverteilung zeigen, daß die Binderegion im Vergleich zur Gesamtoberfläche des Proteins mehr hydrophobe Aminosäuren besitzt. Bei obligatorischen Komplexen ist dies stärker ausgeprägt als bei Komplexen, deren Monomere auch separat in polarem Lösungsmittel existieren [13,23]. Im Vergleich zum Proteininnern sind die Bindungsbereiche allerdings weniger hydrophob. Dennoch spielt der hydrophobe Effekt bei der Bildung von Protein-Protein-Komplexen eine entscheidende Rolle. Die hydrophoben Gruppen der Monomere treffen bei der Komplexbildung aufeinander und werden so von einer polaren in eine unpolare Umgebung überführt. Dabei werden aus der Binderegion Wassermoleküle freigesetzt, was zu einer Entropiezunahme führt und die Komplexbildung begünstigt. Dieser Vorgang ähnelt in gewisser Weise der Proteinfaltung. Es wird jedoch angenommen, daß der hydrophobe Effekt bei der Proteinfaltung eine weitaus wichtigere Rolle spielt als bei der Protein-Protein-Bindung und bei der Faltung sogar die treibende Kraft darstellt [23,27]. Die Zusammensetzung der Bindebereiche stellt einen Kompromiß dar zwischen der Stabilisierung des Komplexes durch den hydrophoben Effekt einerseits und der Vermeidung zu großer hydrophober Bereiche andererseits. Ausgeprägte hydrophobe Stellen wirken sich ungünstig auf die Stabilität der unkomplexierten Monomere in Lösung aus. Daher findet man eher kleine hydrophobe *Patches*, die über den Bindebereich verteilt vorliegen. Der quantitative Beitrag der hydrophoben Wechselwirkung zur Komplexbildung ist noch immer umstritten. Bezogen auf die Bindungsoberfläche werden für den Gewinn an Freier Enthalpie Werte von $105\text{-}210 \text{ J}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ diskutiert [33,34].

Außer den hydrophoben Wechselwirkungen sind auch elektrostatische Wechselwirkungen bei der Komplexbildung beteiligt. In den Komplexen liegt ein hohes Maß an elektrostatischer Komplementarität vor: Regionen positiven elektrostatischen Potentials des einen Proteins treffen bei dem Komplexpartner auf Regionen mit negativem Potential [28]. In der Binderegion befinden sich zwar mehr polare und geladene Aminosäuren als im Proteininnern, jedoch weniger als an der restlichen Proteinoberfläche [13,23]. Dennoch ähneln Protein-Protein-Binderegionen stärker der restlichen Proteinoberfläche als dem Proteininnern. Zwischen den polaren Gruppen der Proteinpartner können sich intermolekulare Wasserstoffbrücken ausbilden, an denen auch geladene Gruppen beteiligt sein

können. Die Anzahl der intermolekularen Wasserstoffbrücken steigt zwar mit der Größe der Binderegion (ca. eine Wasserstoffbrücke pro 170 \AA^2), je nach Komplextyp kann ihre Gesamtzahl jedoch sehr stark variieren (z.B. zwischen 1-34 bei dem Datensatz von Conte *et al.*) [17]. Neben Wasserstoffbrückenbindungen tragen auch intermolekulare Salzbrücken zur Stabilisierung des Komplexes bei, die sich zwischen zwei entgegengesetzt geladenen Gruppen ausbilden. Nussinov *et al.* [26] konnten zeigen, daß bei Protein-Protein-Komplexen in der Binderegion im Durchschnitt zwei Salzbrücken ausgebildet werden. In Monomeren sind polare und geladene Gruppen, die sich an der Proteinoberfläche befinden, normalerweise durch Wassermoleküle solvatisiert. Werden diese Gruppen bei der Komplexbildung in das Innere des Bindebereichs überführt, muß ein erhebliches Maß an Energie für die Desolvatisierung aufgebracht werden. Kann dieser Prozeß nicht durch günstige Wechselwirkungen mit entgegengesetzten Ladungsträgern oder durch die Bildung eines dichten Netzwerkes von Wasserstoffbrücken in der Binderegion kompensiert werden, führt dies zur Destabilisierung des Komplexes [27,29]. Durch den hohen Energieaufwand für die Desolvatisierung scheint der relative Beitrag elektrostatischer Wechselwirkungen zur Bindungsenergie im Vergleich zu den hydrophoben Wechselwirkungen eher gering zu sein. Für die Spezifität der Komplexbildung scheinen polare Wechselwirkungen allerdings eine entscheidende Rolle zu spielen, da die elektrostatische Komplementarität nur für ganz bestimmte Orientierungen und Komplexpartner erfüllt werden kann [27].

Die Kenntnis über die oben besprochenen Faktoren, welche die Proteinkomplexbildung beeinflussen, konnte bereits erfolgreich zur Vorhersage möglicher Binderegionen in Monomeren [35-37] wie auch im Bereich des Protein-Dockings [38-40] angewendet werden. Weitere Untersuchungen heben die Bedeutung bestimmter Aminosäuren oder die Rolle von Wassermolekülen in der Binderegion hervor. Bogan und Thorn [41] beispielsweise identifizierten durch Mutationsexperimente sogenannte *Hot Spots* im Zentrum der Binderegion, die einen wesentlichen Beitrag zur Freien Bindungsenthalpie liefern. Sie bestehen aus einem kleinen Cluster ausgewählter Aminosäuren, die an einen räumlich benachbarten *Hot Spot* des Komplexpartners angrenzen und vom Solvens abgeschirmt sind. Während bei den *Hot Spots* der explizite Ausschluß von Wassermolekülen hervorgehoben wird, gibt es auch Beispiele, bei denen Wasser zwischen den Proteingrenzflächen eine zentrale Rolle spielt. Die Bedeutung von Wassermolekülen in der Binderegion ist zentraler Bestandteil dieser Arbeit und wird im folgenden Kapitel erläutert.

2.2 Wassermoleküle in der Binderegion von Proteinkomplexen

Wasser stellt durch seine strukturellen und physikalischen Eigenschaften für fast alle biologischen Prozesse einen essentiellen Baustein dar. Struktur, Dynamik und Funktion von Proteinen beruhen sehr stark auf Wechselwirkungen mit dem Lösungsmittel Wasser. So tragen Wassermoleküle zur Stabilität der Proteinkonformation bei oder sind an enzymkatalysierten Reaktionen beteiligt. Auch bei der Protein-Protein-Erkennung spielen Wassermoleküle eine wichtige Rolle. Untersuchungen von Kristallstrukturen zeigen, daß sich im Durchschnitt 18-20 Wassermoleküle (ca. ein Wassermolekül pro 100 \AA^2) innerhalb der Binderegion (*Interface*) von Protein-Protein-Komplexen befinden [17,26]. Die Anzahl beteiligter Wassermoleküle im *Interface* ist sehr heterogen und liegt im Datensatz von Nussinov *et al.* [26] zwischen 3 und 50. Janin [7] spricht daher von *Dry Interfaces*, wenn Wassermoleküle nur an den Rändern um die zentrale wasserfreie Bindestelle vorkommen, und von *Wet Interfaces*, wenn die Binderegion viele wassergefüllte Hohlräume (*Cavities*) aufweist. Die Anzahl der tatsächlich in der Binderegion enthaltenen Wassermoleküle ist vermutlich größer als die der beobachteten Kristallwasser, da es noch immer schwierig ist, Solvenspositionen in Kristallstrukturen adäquat vorherzusagen.

Poljak *et al.* [42] konnten für Antigen-Antikörper-Komplexe zeigen, daß Wassermoleküle erheblich zur Komplementarität der Antigen-Antikörper-Bindestelle beitragen und integraler Bestandteil dieser Bindestelle sind. Sie füllen destabilisierende Hohlräume in der Binderegion des Komplexes aus und vernetzen die Komplexpartner mit Wasserstoffbrücken. Ein einziges Wassermolekül verbindet im Mittel vier Atome der Komplexpartner miteinander [26]. Dabei dienen Wassermoleküle als eine Art „Vermittler“, die gleichartig polarisierte Gruppen (Donatoren oder Akzeptoren) miteinander verbrücken, die ansonsten keine direkte Wasserstoffbrücke ausbilden können. Es wird vermutet, daß durch Wasser vermittelte polare Bindungen häufiger sind als direkte Wasserstoffbrücken zwischen den Komplexpartnern [17]. Im Vergleich zum Innern eines Proteins werden in der Binderegion mehr Wassermoleküle beobachtet, was auf die stärker hydrophobe Umgebung im Innern eines Protein zurückgeführt werden kann. In der Packungsdichte unterscheidet sich die Binderegion nicht vom Innern des Proteins [17], da die Hohlräume im *Interface* durch Wassermoleküle ausgefüllt sind und so eine dichte Packung erreicht werden kann.

Neben der Spezifität tragen die Wassermoleküle in der Binderegion auch zur Stabilität des Komplexes bei. Bhat *et al.* [43] konnten an Antigen-Antikörper-Komplexen zeigen, daß die bei Antikörpern hauptsächlich durch Enthalpie getriebene Bindung durch die im Bindebereich vorhandenen Wassermoleküle stabilisiert wird. Wassermoleküle tragen erheblich zum enthalpischen Anteil der Bindestärke von Proteinkomplexen bei und können bis zu 25% der Gesamtaffinität darstellen [44,45]. Mutationsexperimente zeigen, daß die durch Mutationen eingeführten Komplementaritätsverluste in der Binderegion eher durch zusätzlich eingebundene Wassermoleküle ausgeglichen und stabilisiert werden als durch Konformationsänderungen der Proteine [46]. Obwohl Wassermoleküle stark zur Affinität und Spezifität von Proteinkomplexen beitragen, scheinen sie nicht auf ihren Positionen in der Binderegion „gefangen“ zu sein. NMR-Experimente zeigen Verweilzeiten zwischen 10^{-2} Sekunden für Wasser in Hohlräumen innerhalb des *Interfaces* bis 10^{-10} Sekunden und schneller für Stellen direkt an der Proteinoberfläche [7,9]. Durch den schnellen Austausch von Wassermolekülen in Hohlräumen sind diese immer mit Wassermolekülen besetzt, und die dichte Packung der Binderegion kann aufrechterhalten werden.

Die dargelegten Untersuchungen zeigen die bedeutende Rolle von Wassermolekülen bei der Bildung von Proteinkomplexen. Durch Vergleiche von Kristallstrukturen der unkomplexierten Proteinpartner mit der Komplexstruktur können Kristallwassermoleküle in der Binderegion identifiziert werden, die sowohl im Komplex als auch in den Monomerstrukturen vorliegen. Auf Basis solcher Ergebnisse wurden Verfahren entwickelt, diese „konservierten“ Kristallwassermoleküle, die an der Bindung teilnehmen, in den Monomerstrukturen vorherzusagen. Diese Methoden beruhen auf genetischen Algorithmen [47] oder multilinearer Regression [48], die Deskriptoren wie Protein-Wasser-Kontaktfläche oder kristallographische Temperaturfaktoren für die Vorhersage verwenden. Gerade für *Docking*-Experimente ist die Kenntnis über konservierte Wassermoleküle sehr nützlich, da diese dann von vornherein bei dem *Docking*-Vorgang berücksichtigt werden können. Es existiert auch schon ein Ansatz, bei dem einzelne Wassermoleküle während des Dockingvorgangs in die Binderegion integriert werden [49]. Zusätzlich gibt es Hinweise dafür, daß die Berücksichtigung von Information über Wassermoleküle in der Binderegion bei QSAR-Studien (*Quantitative Structure-Activity Relationships*) zu besseren Ergebnissen führt, was bei der Vorhersage relativer Bindeaffinitäten unterschiedlicher Inhibitoren an das Enzym Glycogenphosphorylase gezeigt werden konnte [50].

Häufig sind in Proteinkristallstrukturen keine Koordinaten von Wassermolekülen verfügbar. Ein Weg zum Auffinden möglicher Wasserpositionen ist die Suche nach Hohlräumen in den Proteinstrukturen. Zur Identifizierung solcher Hohlräume in Proteinen existieren inzwischen zahlreiche Algorithmen und Programme [51-56]. Bei den meisten Algorithmen handelt es sich um geometrische Ansätze, die mit Hilfe von Kugeln die Kristallstruktur nach Hohlräumen oder Kanälen absuchen. Häufig wird in den Rechenverfahren der abzusuchende Raum durch ein dreidimensionales Gitter repräsentiert [51,52] oder das Molekül durch das Konzept molekularer Oberflächen (siehe Kapitel 2.4) dargestellt [56]. Ein Nachteil der Verfahren ist, daß nicht zwischen Hohlräumen im Innern des Proteins und solchen in der Binderegion unterschieden wird. Nach der Berechnung der Hohlräume läßt sich einfach feststellen, ob und welche Kristallwassermoleküle sich darin befinden. Die aufgeführten Verfahren sind allerdings nicht dafür optimiert, unbesetzte Hohlräume mit Wassermolekülen aufzufüllen. Wenn die Kristallstruktur keine Koordinaten von Wassermolekülen enthält, kann weder die Anzahl noch die Position der Wassermoleküle in den berechneten Hohlräumen vorhergesagt werden. Mit dem von Zhang und Hermans entwickelten Programm *DOWSER* ist eine Berechnung potentieller Wasserpositionen in Proteinhohlräumen möglich [57]. Sie postulieren, daß ein Hohlraum mit einem Wassermolekül besetzt werden kann, wenn die Wechselwirkungsenergie zwischen Protein und Wassermolekül einen Wert von $-50 \text{ kJ}\cdot\text{mol}^{-1}$ nicht überschreitet. Basierend auf diesen Energieberechnungen können Wassermoleküle mit *DOWSER* in leere Hohlräume und Taschen von Proteinen plaziert werden. Das Programm eignet sich allerdings nicht sehr gut zum Füllen großer Hohlräume, da Wassermoleküle immer nur am Rand eines Hohlraumes positioniert werden, so daß das Verfahren bei großen Hohlräumen mehrfach wiederholt werden muß.

Untersuchungen von Hohlräumen und deren struktureller und chemischer Umgebung beschränkten sich bisher auf im Proteininnern verborgene Hohlräume globulärer Monomere [58-60]. Hohlräume in der Binderegion von Protein-Ligand- [61] und Protein-Protein-Komplexen sind hingegen wenig charakterisiert. Dies spiegelt sich auch bei statistischen Analysen der Binderegionen wider, wo ausschließlich die in der Kristallstruktur vorhandenen Wassermoleküle berücksichtigt werden, anstatt auch leere Hohlräume mit Wassermolekülen aufzufüllen und die tatsächlich in der Binderegion vorhandene Wasseranzahl abzuschätzen [17,26]. Dies bildete die Motivation der

vorliegenden Arbeit zur Entwicklung eines schnellen und automatisierten Verfahrens, mit dem Hohlräume in der Binderegion von Protein-Protein-Komplexen erkannt und die optimale Anzahl der darin enthaltenen Wassermoleküle vorhergesagt werden können (siehe Kapitel 3).

2.3 Vorhersage der Affinität von Proteinkomplexen

Der molekulare Erkennungsprozeß zweier Proteine, der zur Ausbildung eines größeren Proteinkomplexes führt, beinhaltet zwei essentielle Faktoren: *Spezifität* und *Affinität*. Spezifität macht die molekulare Erkennung zu einem Instrument für die Regulierung komplexer Zellprozesse, Affinität beschreibt die Stärke der Wechselwirkungen zwischen den Komplexpartnern. Biologische Prozesse beruhen im wesentlichen darauf, daß Proteine ein unterschiedliches Maß an Spezifität oder Affinität zueinander aufweisen. Die Stärke von Proteinkomplexen variiert je nach physiologischer Funktion und wird durch die Änderung der Freien Enthalpie während der Komplexbildung quantitativ beschrieben [62,63]. Die Dissoziationskonstante des Komplexes ist mit der Freien Enthalpie verknüpft und kann experimentell durch isotherme Titrationskalorimetrie [64,65] oder Fluoreszenz-Spektroskopie [66] bestimmt werden. Proteinkomplexe weisen einen großen Bereich von Dissoziationskonstanten auf (10^{-5} - 10^{-13} M); die experimentellen Bestimmungsmethoden werden allerdings ab Werten von 10^{-11} M sehr aufwendig und ungenau, da mit extrem geringen Proteinkonzentrationen gearbeitet werden muß.

Die experimentell bestimmten Dissoziationskonstanten können mit den strukturellen und physikalischen Eigenschaften der Komplexe in Beziehung gesetzt werden, um ein besseres Verständnis von Proteinwechselwirkungen zu bekommen und theoretisch Ansätze zur Vorhersage von Bindungsaffinitäten bei Proteinkomplexen zu entwickeln. Gerade in den Bereichen *Docking*, *Protein Engineering* und *Drug Design* sind theoretische Verfahren zur Unterscheidung schwacher und starker Komplexwechselwirkungen notwendig. Hierbei muß eine große Anzahl möglicher Komplexe getestet werden, was nur durch Computersimulationen mit vertretbaren Kosten bewerkstelligt werden kann. Zur Vorhersage von Bindungsaffinitäten existieren mehrere empirische Ansätze. Eine recht genaue Methode beruht auf Molekulardynamiksimulationen mit sogenannten *Free-Energy-Perturbation*-Berechnungen (FEP) [67]. Der Simulationsaufwand ist allerdings erheblich

und das Verfahren nicht schnell genug, um die Affinitäten aller Strukturen großer Datenbanken zu einem bestimmten Zielprotein zu testen. Eine schnellere Methode stellt die Berechnung von wissensbasierten Paarpotentialen (*Potential of Mean Force, PMF*) dar [68]. Hierbei handelt es sich um ein statistisches Verfahren, bei dem die Häufigkeiten bestimmter struktureller Faktoren innerhalb eines Testdatensatzes, beispielsweise das Auftreten zweier unterschiedlicher Atomtypen in einem definierten Abstand zueinander, in Paarpotentiale umgewandelt werden. Die Verknüpfung der aufgestellten Potentiale mit den Bindungskonstanten des Testdatensatzes zu einer sogenannten *Scoring*-Funktion kann dann die Vorhersage von Affinitäten strukturell ähnlicher Komplexe ermöglichen. Andere Methoden verwenden ebenfalls *Scoring*-Funktionen [10] zur Bewertung der Affinität von Proteinkomplexen. Sie basieren meist auf einem linearen Regressionsansatz, bei dem die Freie Bindungsenthalpie als Summe physikalisch sinnvoller Teilbeiträge (Deskriptoren) ausgedrückt wird. Als Deskriptoren werden beispielsweise hydrophobe oder hydrophile Oberflächenanteile, Wasserstoffbrücken oder die Anzahl durch die Komplexbildung eingefrorener Torsionen verwendet. Die Beiträge der Parameter sind in der *Scoring*-Funktion an die experimentellen Bindungskonstanten eines Testdatensatzes angepaßt. Die Güte der *Scoring*-Funktion hängt sehr stark von der Verlässlichkeit der im Testdatensatz verwendeten Strukturdaten und deren experimentell bestimmten Bindungskonstanten ab. Wie gut die Vorhersage für Komplextypen ist, die nicht durch den Testdatensatz dargestellt werden, ist unklar. Eine *Scoring*-Funktion kann daher nicht einfach auf unterschiedliche Systeme angewendet werden. So existieren *Scoring*-Funktionen für Komplexe aus Proteinen und kleinen Liganden mit Testdatensätzen zwischen 45 und 170 Komplexen [69-72], wie auch für reine Protein-Protein-Komplexe [27,73-75], bei denen die Testdatensätze aus Mangel an experimentellen Daten meist nur aus 20 Komplexen bestehen. Bei der Verwendung von *Scoring*-Funktionen oder der *PMF*-Methode werden Konformationsänderungen der beteiligten Proteine durch die Komplexbildung (*Induced Fit*) meist vollständig vernachlässigt, da ausschließlich die Struktur des Komplexes zur Bewertung der Affinität herangezogen wird. Der Vorteil der beiden Methoden liegt ganz klar in der Geschwindigkeit der Affinitätsvorhersage, so daß eine große Anzahl von Komplexen evaluiert werden kann.

In den zitierten *Scoring*-Funktionen für Protein-Protein-Komplexe bleiben Wassermoleküle in den Binderegionen der Komplexe unberücksichtigt, da meist keine experimentellen Kristallwasserdaten vorliegen. Selbst die Hohlräume zwischen den Komplexpartnern, in denen sich Wassermoleküle aufhalten können, werden in den Bewertungsfunktionen vernachlässigt. Diese Wassermoleküle können allerdings erheblich zur Bindeaffinität beitragen. In Molekulardynamiksimulationen mit *FEP*-Berechnungen konnte gezeigt werden, daß die Integration von Wassermolekülen, die in der Kristallstruktur aufgelöst wurden und sich in der Binderegion befinden, die Vorhersage der Bindeaffinität verbessert [67]. In der vorliegenden Arbeit wurde daher ein Konzept entwickelt, die strukturellen Informationen über mögliche Wasserpositionen in der Binderegion als Deskriptoren in eine *Scoring*-Funktion aufzunehmen (siehe Kapitel 5).

2.4 Molekulare Oberflächen

Die Darstellung von Molekülen beruht je nach Aufgabenstellung auf Modellen unterschiedlicher Komplexität. Reaktionsmechanismen lassen sich häufig durch einfache, zweidimensionale Strichformeln darstellen, während bei der Untersuchung der Wechselwirkungen von Makromolekülen die Beschreibung der dreidimensionalen Struktur notwendig ist. Zur Veranschaulichung solch komplexer Strukturen wurden früher Drahtmodelle verwendet, während heutzutage die Visualisierung der Moleküle mit Hilfe von Computergraphik zum Einsatz kommt. Neben der Information über die Konformation der Moleküle ist bei vielen Untersuchungen auch die Kenntnis der Raumerfüllung des Moleküls von Bedeutung. Moleküle lassen sich zwar durch Elektronendichteverteilungen darstellen [76], die mit zunehmendem Abstand gegen Null streben, sie besitzen jedoch keine physikalisch begrenzte Oberfläche. Aus diesem Grund wird häufig das Konzept „virtueller Moleküloberflächen“ verwendet. Die Modellvorstellung einer Moleküloberfläche erweist sich besonders bei der Beurteilung sterischer Wechselwirkungen zwischen Molekülen, wie z.B. in Proteinkomplexen, als hilfreich, da sie einen Eindruck von Größe, Ausdehnung und topographischen Eigenschaften der betrachteten Moleküle vermittelt. Die Untersuchung der sterischen Komplementarität ist nach der Schlüssel-Schloß-Hypothese von E. Fischer [2] ein wichtiges Kriterium für die Spezifität der Bindung. Des weiteren lassen sich auch ortsabhängige, physikochemische Daten der Moleküle (z.B. Elektrostatik, Lipophilie) farbcodiert auf den Oberflächen darstellen. Diese

Informationen haben sich zur Analyse von Moleküleigenschaften bereits als hilfreicher Ansatz erwiesen [35,38,77-81]. Im folgenden werden kurz unterschiedliche Konzepte molekularer Oberflächen vorgestellt.

2.4.1 Die van-der-Waals-Oberfläche

Der wohl bekannteste Ansatz zur Volumendarstellung von Molekülen ist das auf Corey, Pauling und Koltun zurückgehende Kalottenmodell, das auch als CPK-Darstellung bezeichnet wird [82]. Dabei werden die Atome als hartschalige Kugeln mit definiertem Kugelradius, dem sogenannten *van-der-Waals-Radius*, angesehen. Die van-der-Waals-Radien ergeben sich jeweils aus dem Mindestabstand zweier gleichartiger, nicht-kovalent-gebundener Atome. Die in dieser Arbeit verwendeten Werte von van-der-Waals-Radien beruhen auf dem von Bondi [83] zusammengestellten Datensatz. Im Fall einer chemischen Bindung durchdringen sich die Kugeln, wobei die äußere Hülle der überlappenden Kugeln die van-der-Waals-Oberfläche bildet (Abb. 2.1-A).

2.4.2 Lösungsmittelzugängliche Oberfläche

Bei der Beschreibung von Molekülwechselwirkungen mit Hilfe von Molekülmodellen spielt das umgebende Lösungsmittel eine wichtige Rolle. Es ist daher hilfreich, die für das Lösungsmittel zugängliche Moleküloberfläche bestimmen zu können. Lee und Richards [84] schlugen 1971 erstmals einen Ansatz vor, bei dem ein kugelförmiges Probeteilchen über die van-der-Waals-Oberfläche des Moleküls gerollt wird. Die Mittelpunkte der Probekugeln bilden dann die sog. lösungsmittelzugängliche Oberfläche (*Solvent Accessible Surface*, Abb. 2.1-B). Der Radius der abtastenden Probekugel wird meist von den Dimensionen eines Wassermoleküls, dem wichtigsten Lösungsmittel in biologischen Systemen, abgeleitet und liegt im Bereich von 1.4 Å. Kurze Zeit später entwickelte Richards [85] die Definition der *Solvent Excluded Surface*, die häufig auch als *Connolly-Oberfläche* bezeichnet wird. Sie wird aus den konvexen Kontaktbereichen der van-der-Waals-Oberfläche mit der Probekugel (*Contact Surface*) und den interpolierten konkaven Bereichen zwischen den Atomen (*Reentrant Surface*) gebildet (Abb. 2.1-C). Das von der Oberfläche eingeschlossene Volumen stellt somit den Raum dar, der vom umgebenden Lösungsmittel nicht erreicht werden kann.

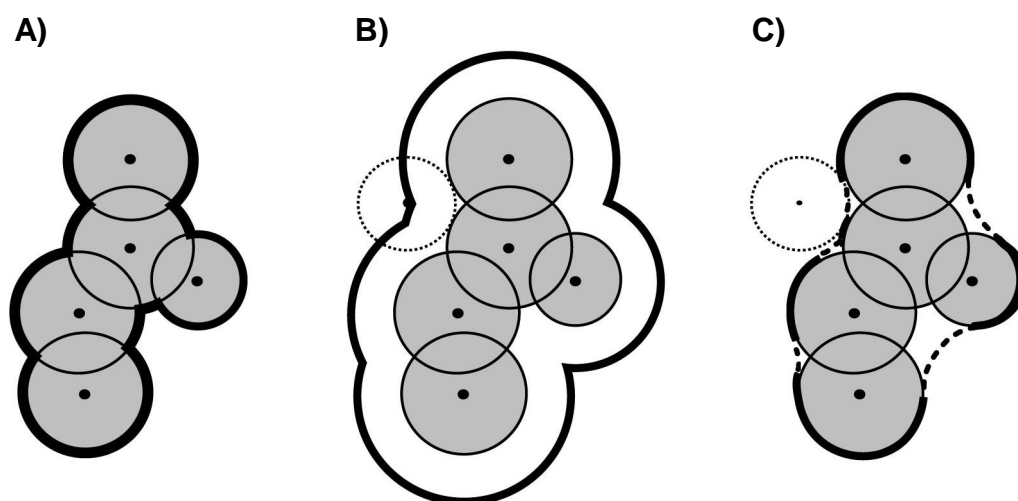


Abb. 2.1: Schematische Darstellung der unterschiedlichen Definitionen molekularer Oberflächen (umhüllende Linie). Atome werden durch graue Kugeln repräsentiert, Probesteilchen zur Generierung der Oberfläche durch gestrichelte Kugeln. **A)** van-der-Waals-Oberfläche **B)** *Solvent Accessible Surface* (Lee & Richards) **C)** Connolly-Oberfläche, bestehend aus *Contact Surface* (—) und *Reentrant Surface* (- - -).

Connolly entwickelte mit dem Programm *MS* erstmals einen Algorithmus, mit dem die Berechnung und Darstellung der *Solvent Excluded Surface* als Punktoberfläche am Computer möglich war [86,87]. Da sich die Darstellung einer Oberfläche als Polygonnetz für geometrische Untersuchungen und zur Visualisierung jedoch besser eignet, wurde von Heiden [88] eine Methode zur Triangulation der Punkte vorgeschlagen, welche die Punktoberfläche in ein Dreiecksnetz überführt. Neben den skizzierten Methoden existieren inzwischen auch schnelle analytische Methoden, mit denen direkt eine triangulierte Oberfläche erzeugt werden kann [89,90]. In der vorliegenden Arbeit wird eine von Keil [35] entwickelte, gitterbasierte Methode zur Berechnung der molekularen Oberfläche verwendet. Vorteile sind die hohe Geschwindigkeit sowie die geringe Fehleranfälligkeit des Verfahrens. Bei dem von Keil entwickelten Programm *fumee* wird das Molekül in ein dreidimensionales Gitter gelegt, auf dem die Abstände zum jeweils nächstgelegenen Atom abgelegt sind. Mit Hilfe des *Marching-Cube*-Algorithmus [91] wird anschließend eine Isofläche durch Orte gleicher Abstandswerte generiert, die der molekularen Oberfläche entspricht. Mit diesem Verfahren lassen sich sowohl die *Solvent Accessible Surface* als auch eine sehr gute Approximation der Connolly-Oberfläche erzeugen. Für alle folgenden Untersuchungen an Oberflächen in dieser Arbeit wird ausschließlich die so approximierte Connolly-Oberfläche verwendet.

2.4.3 *Separating Surface*

Um die Analyse der Wechselwirkungen zwischen Komplexpartnern zu erleichtern, wurden Ansätze entwickelt, die Kontaktregionen beider Komplexpartner durch eine einzige Kontaktoberfläche darzustellen [92-94]. Der Vorteil einer solchen Oberfläche liegt darin, daß die topographischen und physikochemischen Eigenschaften beider Komplexpartner gleichermaßen auf ihr abgebildet und verglichen werden können. Gabdoulline und Wade [93,94] transformieren die Daten der dreidimensionalen Oberfläche anschließend in eine Ebene, um eine zweidimensionale Darstellung der Ergebnisse zu ermöglichen. Von Keil [95] wurde die sogenannte *Separating Surface* entwickelt (Abb. 2.2). Sie ist durch einen Satz von Punkten definiert, die sich genau zwischen den beiden van-der-Waals-Oberflächen in der Binderegion der Komplexpartner befinden. Damit werden Größe und Aussehen der *Separating Surface* durch die Bindestellen der Komplexpartner bestimmt. Die farbige Darstellung der Atomabstände zwischen den Komplexpartnern auf der Oberfläche gibt beispielsweise Aufschluß über mögliche Hohlräume oder Atomüberlappungen in der Binderegion. Daneben lassen sich auf der *Separating Surface* auch die physikochemischen Eigenschaften beider Moleküle in der Binderegion darstellen. Das bietet einen großen Vorteil, da bisher die Identifikation von Bereichen attraktiver Wechselwirkungen zwischen den Komplexpartnern innerhalb der Binderegion mit Hilfe von separaten Oberflächenmodellen nur qualitativ durch die getrennte Untersuchung beider Komplexoberflächen erfolgen konnte. Mit Hilfe des Konzeptes der *Separating Surface* können nun Ansätze verfolgt werden, die eine quantitative Aussage über die Komplementarität der Bindeflächen bezüglich der physikochemischen Moleküleigenschaften ermöglichen. Dazu eignet sich die Berechnung von Komplementaritäts- bzw. Similaritätsindizes, wie beispielsweise des Carbo- [96] oder Hodgkin-Index [97], auf der *Separating Surface*.

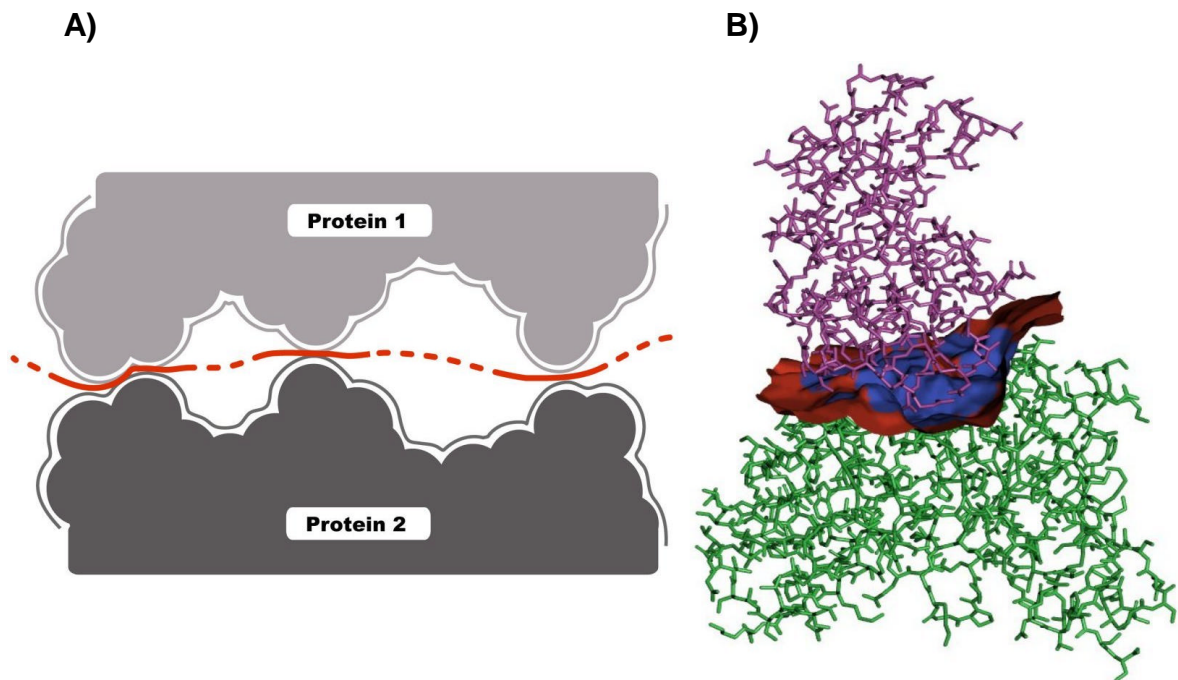


Abb. 2.2: **A)** Schematische Darstellung der *Separating Surface* (rot) in der Binderegion zweier Komplexpartner (grau). Oberflächenteile, die einen bestimmten *Cutoff*-Abstand zu den Komplexpartnern überschreiten sind unterbrochen dargestellt. **B)** *Separating Surface* der Kristallstruktur (1vfb) des Antikörperfragments D1.3 (grün) mit seinem Antigen Lysozym (magenta). Blaue Oberflächenteile sind weniger als 1.3 Å von den van-der-Waals-Oberflächen beider Proteine entfernt und beschreiben direkte Protein-Protein Kontakte. Rote Anteile der Oberfläche sind weiter als 1.3 Å von den Proteinen entfernt. Sie bilden die Randbereiche der *Separating Surface*, die mit dem umgebenden Lösungsmittel in Kontakt stehen. Sofern sie sich im Innern der Binderegion befinden, weisen sie auf abgeschlossene Hohlräume hin.

2.5 Darstellung physikochemischer Daten auf molekularen Oberflächen

Molekulare Oberflächen geben nicht nur Aufschluß über Form und Größe von Molekülen, sondern eignen sich auch sehr gut als Projektionsfläche für ortsabhängige, physikochemische Daten, wie z.B. Elektrostatik oder Lipophilie. Dafür werden die zu untersuchenden Moleküleigenschaften an jedem Oberflächenpunkt berechnet und anschließend farbcodiert auf der Oberfläche dargestellt. Die in der vorliegenden Arbeit verwendeten Moleküleigenschaften zur Analyse von Proteinkomplexen werden in den folgenden Kapiteln besprochen.

2.5.1 Elektrostatisches Potential

Bei der Beschreibung intermolekularer Wechselwirkungen von Molekülkomplexen ist die Untersuchung der elektrostatischen Komplementarität in der Binderegion von Bedeutung, da sie zur Bindungsstärke des Komplexes beiträgt. Neben den Komplexpartnern muß auch das Lösungsmittel bei der Berechnung des elektrostatischen Potentials j berücksichtigt werden, da biochemische Reaktionen überwiegend in wäßriger Lösung ablaufen. Die Unterschiede zwischen Solvat und Lösungsmittel zeigen sich beispielsweise in den unterschiedlichen Dielektrizitätskonstanten. Für das Innere von Proteinen werden Werte für die relative Dielektrizitätskonstante von $\epsilon_r = 2-10$ diskutiert [98], während der Wert für Wasser bei $\epsilon_r = 80$ liegt. Ein Ansatz zur Berechnung des elektrostatischen Potentials, der die unterschiedliche Polarisierbarkeit von Solvat und Lösungsmittel berücksichtigt, ist die Poisson-Boltzmann-Gleichung. Dabei wird das Lösungsmittel als Kontinuum um eine Kavität modelliert, die durch das gelöste Molekül gebildet wird [99,100]. Da für komplexe Makromoleküle eine analytische Lösung nicht möglich ist, werden iterative Finite-Differenzen-Algorithmen zur Lösung der Gleichung verwendet [101,102]. Der Rechenaufwand steigt dabei sehr schnell mit der Größe des Systems. In der vorliegenden Arbeit wird daher ein vereinfachtes Modell zur Berechnung des elektrostatischen Potentials gewählt, das auf dem Coulomb-Gesetz (Gl. 2.1) beruht:

$$j(\vec{r}) = \frac{1}{4\pi\epsilon} \cdot \sum_{i=1}^N \frac{q_i}{d_i} \quad (2.1)$$

mit N : Anzahl der Atome des Moleküls,
 q_i : Partialladung des Atoms i ,
 $\epsilon = \epsilon_r \cdot \epsilon_0$: Dielektrizitätskonstante (DE), mit relativer DE ϵ_r und DE im Vakuum ϵ_0 ,
 d_i : Abstand des Atoms i zum Ort der Potentialberechnung \vec{r} mit $d_i = |\vec{r} - \vec{r}_i|$.

Unter Verwendung punktförmiger Partialladungen auf den Atomschwerpunkten kann mit dem Coulomb-Gesetz das elektrostatische Potential in einem homogenen Medium bestimmt werden. Die Partialladungen lassen sich für kleinere Moleküle mit semiempirischen oder *ab-initio*-Programmen berechnen. Für biologische Makromoleküle ist die Berechnung aller Partialladungen wegen des enormen Rechenaufwands jedoch nicht

sinnvoll. Stattdessen werden die Partialladungen für Makromoleküle sogenannten Kraftfeldern entnommen, in denen die Partialladungen für unterschiedliche Molekülfragmente bestimmt und tabelliert wurden. Ein für Proteine und Nukleinsäuren häufig verwendetes Kraftfeld ist das auch in dieser Arbeit verwendete *CHARMM*-Kraftfeld [103]. Die Dielektrizitätskonstante besitzt im Coulomb-Ansatz einen über den gesamten Raum konstanten Wert, obwohl diese wie oben dargelegt ortsabhängig ist. Für die Approximation der Polarisierbarkeit des Lösungsmittels, und der damit verbundenen stärkeren Abschirmung der Ladungen, wird die Abstandsfunktion des Coulomb-Ansatzes durch einen Zusatzterm modifiziert, der das elektrostatische Potential skaliert:

$$\mathbf{j}(\vec{r}) = \begin{cases} \frac{1}{4\pi\epsilon} \cdot \sum_{i=1}^N \frac{q_i}{d_i} \cdot \left[1 - \left(\frac{d_i}{d_0} \right)^2 \right]^2 & \text{für } d_i \leq d_0 \\ 0 & \text{für } d_i > d_0 \end{cases} \quad (2.2)$$

mit N : Anzahl der Atome des Moleküls,

q_i : Partialladung des Atoms i ,

d_0 : *Cutoff*-Radius,

ϵ : Dielektrizitätskonstante mit $\epsilon = \epsilon_r \cdot \epsilon_0$,

d_i : Abstand des Atoms i zum Ort der Potentialberechnung \vec{r} mit $d_i = |\vec{r} - \vec{r}_i|$.

Bei diesem *Shifted-Force*-Ansatz (Gl. 2.2) sinkt das elektrostatische Potential schneller ab als bei der einfachen $1/d$ -Abhängigkeit des Coulomb-Terms und erreicht bei einem bestimmten *Cutoff*-Radius den Wert Null (Abb. 2.3). Der modifizierte Coulomb-Ansatz eignet sich für die schnelle Berechnung des elektrostatischen Potentials auf der molekularen Oberfläche, also im Kontaktbereich des Moleküls zum Solvens, und wurde bereits für die qualitative Analyse von Moleküleigenschaften verwendet [35,38,81]. Dabei zeigte sich eine gute Übereinstimmung der Werte des elektrostatischen Potentials mit dem sehr viel rechenintensiveren Poisson-Boltzmann-Ansatz. In dieser Arbeit wird ein *Cutoff* von 11 Å und eine relative Dielektrizitätskonstante von Eins verwendet. Darüber hinaus kann der Ansatz auch zum quantitativen Vergleich der relativen Größen des elektrostatischen Potentials unterschiedlicher Moleküle herangezogen werden [35].

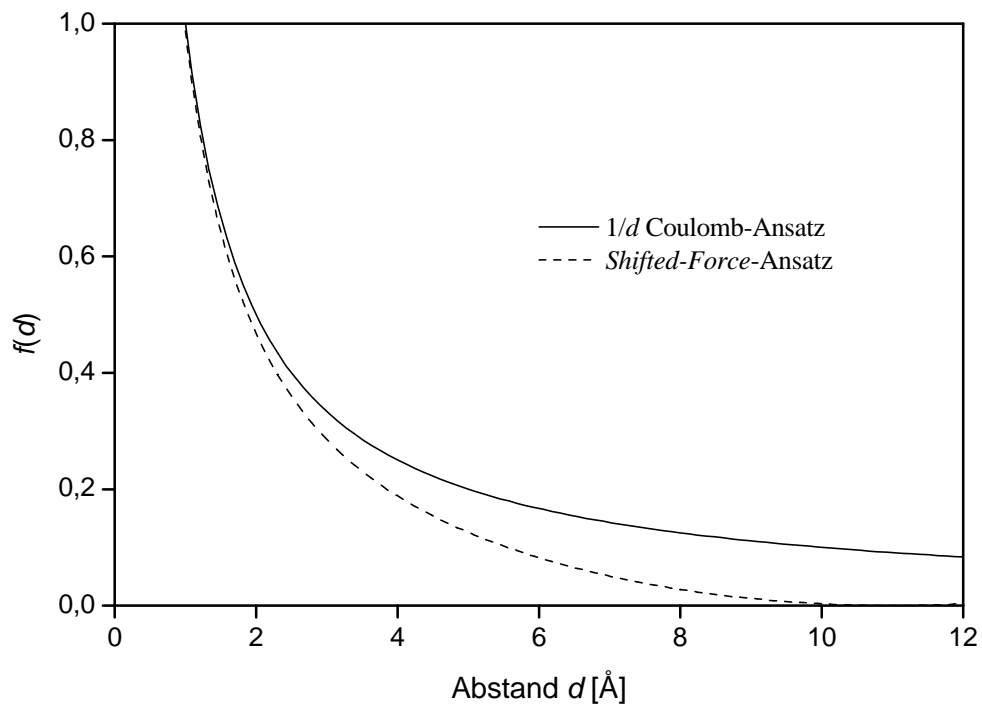


Abb. 2.3: Die Abstandsabhängigkeit des Coulomb-Ansatzes (Gl. 2.1) und des *Shifted-Force*-Ansatzes (Gl. 2.2) wird durch die jeweilige Abstandsfunction $f(d)$ wiedergegeben. Bei dem *Cutoff*-Radius von 11 Å erreicht die Funktion beim *Shifted-Force*-Ansatz den Wert Null.

2.5.2 Lokale Lipophilie

Neben den elektrostatischen Wechselwirkungen spielen bei der Bildung von Proteinkomplexen auch hydrophobe Wechselwirkungen eine wichtige Rolle. Zur Beschreibung der Komplexbildung ist daher eine Quantifizierung lokaler hydrophober Eigenschaften der Komplexpartner von großem Interesse. Ein Ansatz dafür bietet die Bestimmung der Änderung der Freien Enthalpie des Transfers eines Moleküls von einer polaren in eine unpolare Phase. Der Verteilungskoeffizient P beschreibt dabei das Verteilungsgleichgewicht des Moleküls in den beiden nicht mischbaren Flüssigkeiten und kann als quantitatives Maß für die Hydrophobie eines Moleküls dienen. Als Referenzsystem für nicht mischbare Flüssigkeiten wird meist Wasser/n-Oktanol verwendet. Auf Basis der von Fujita *et al.* [104] postulierten additiv-konstitutiven Natur des logarithmierten Verteilungskoeffizienten $\log P$ wurden Ansätze [105] entwickelt, Moleküle in Fragmente mit definierten Lipophilie-Partialwerten f_i zu zerlegen, aus denen sich dann der $\log P$ -Wert berechnen läßt:

$$\log P = \sum_i f_i . \quad (2.3)$$

Lipophilie-Partialwerte können sowohl für Molekülfragmente als auch für einzelne Atome ermittelt werden, indem sie mit Hilfe statistischer Regressionsverfahren an eine Vielzahl experimenteller Verteilungskoeffizienten unterschiedlicher Moleküle angepaßt werden. In der vorliegenden Arbeit wurde der von Ghose und Crippen [106,107] vorgeschlagene und von Viswanadhan *et al.* [108] bzw. Keil [77] erweiterte Parametersatz von Lipophilie-Partialwerten verwendet. Ghose und Crippen definierten 110 Atomtypen, welche die Lipophilie-Partialwerte von Atomen unterschiedlicher funktioneller Gruppen unter Berücksichtigung ihrer strukturellen Umgebung beschreiben. Das System wurde von Viswanadhan *et al.* auf 120 Atomtypen erweitert, und Keil ergänzte anschließend Parameter für die in Proteinen vorkommenden geladenen Aminosäuren Arginin, Lysin, Asparagin- und Glutaminsäure.

Die atomaren Lipophilie-Partialwerte können neben der Berechnung des $\log P$ -Wertes auch zur Darstellung der lokalen Lipophilie des Moleküls auf der molekularen Oberfläche verwendet werden. Audry *et al.* [109] schlugen dafür das *Molecular Lipophilicity Potential* (*MLP*) vor, das nicht als Potential im physikalischen Sinne aufgefaßt werden kann, sondern eine abstandsgewichtete Summierung der Lipophilie-Partialwerte f_i an einem bestimmten Punkt r darstellt:

$$MLP(r) = \sum_i f_i \cdot \frac{1}{1 + d_i/d_0} = \sum_i f_i \cdot g(d_i) \quad (2.4)$$

wobei d_i der Abstand des Fragments i zum Oberflächenpunkt r ist (mit $d_0 = 1 \text{ \AA}$) und $g(d_i)$ die Abstandsfunktion beschreibt.

Daneben wurden noch weitere Abstandsfunktionen zur Berechnung des *MLP* entwickelt, die sich für kleine Moleküle bewährt haben [110-112]. Die in dieser Arbeit verwendete Abstandsfunktion von Heiden und Möckel [113] eignet sich sehr gut zur Untersuchung der Lipophilie an den molekularen Oberflächen von Proteinen. Da bei der Hydrophobie keine langreichweitigen Wechselwirkungen auftreten, werden bei der Heiden-Möckel-Funktion

$g_{HM}(d_i)$ im Proteininnern liegende Atome sehr viel schwächer gewichtet als Atome nahe der Oberfläche:

$$g_{HM}(d_i) = \frac{1 + e^{-(a \cdot b)}}{1 + e^{a \left(\frac{d_i}{d_0} - b \right)}} \quad (2.5)$$

wobei die Konstanten a und b Lage und Breite des Funktionsabfalls angeben (Abb. 2.4).

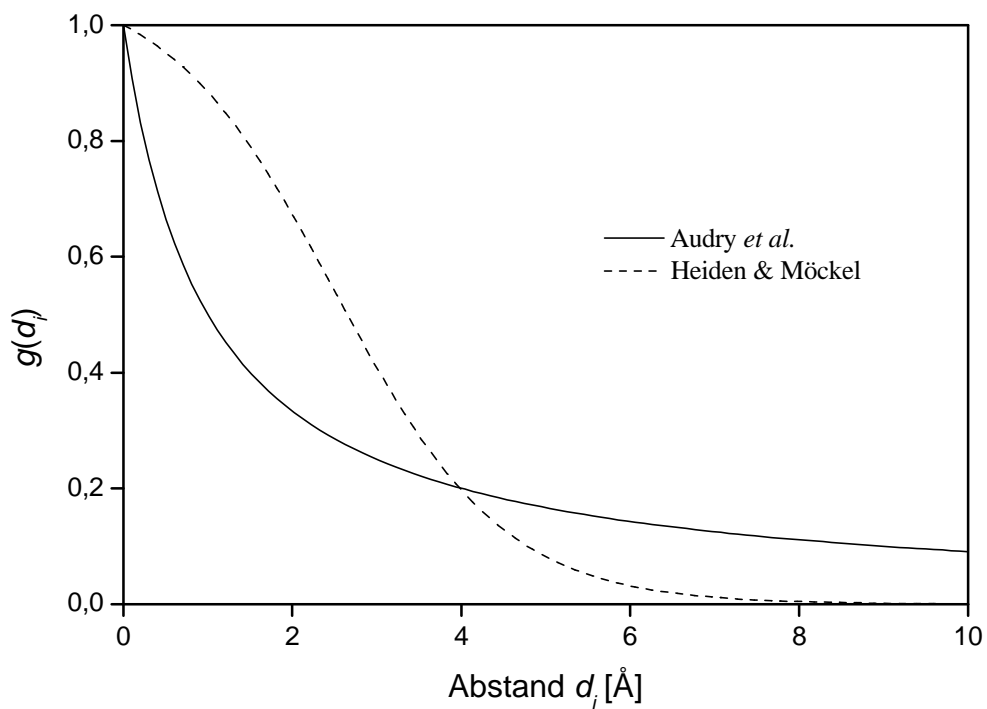


Abb. 2.4: Die Abstandsabhängigkeit $g(d_i)$ zur Berechnung der lokalen Lipophilie nach Audry *et al.* (Gl. 2.4) bzw. nach Heiden und Möckel (Gl. 2.5, mit $a = 1.0$ und $b = 2.5$). Im Gegensatz zu der Funktion von Audry *et al.* wird bei dem Heiden-Möckel-Ansatz die lokale Nachbarschaft von Atomen zum Ort der Potentialberechnung stärker berücksichtigt, während der Einfluß weiter entfernter Atome schnell abklingt.

Bei der Berechnung des ortsabhängigen MLP wird bei dem Ansatz von Heiden und Möckel ein zusätzlicher Normierungsfaktor eingeführt, damit die Werte nur noch in der Größenordnung der Lipophilie-Partialwerte f_i liegen. Das MLP ergibt sich nach Heiden und Möckel demnach für jeden Oberflächenpunkt als:

$$MLP_{HM}(r) = \frac{\sum_i f_i \cdot g_{HM}(d_i)}{\sum_i g_{HM}(d_i)}. \quad (2.6)$$

2.5.3 Wasserstoffakzeptoren und -donatoren

Ein wichtiger Bindungstyp in biologischen Systemen ist die Wasserstoffbrückenbindung. Dabei handelt es sich um eine gerichtete, elektrostatische Wechselwirkung zwischen einer schwach aziden Donorgruppe (D-H) und einem Akzeptoratom (A). Der Abstand des Wasserstoffatoms der Donorgruppe zum Akzeptoratom ist bei der Wasserstoffbrückenbindung kürzer als der erwartete van-der-Waals Kontaktabstand der beiden Atome. Im optimalen Fall liegt eine lineare D-H...A Konfiguration vor, wobei die Entfernung der beiden elektronegativen Atome D...A meist zwischen 2.7 und 3.1 Å liegt. Die Atome D und A sind elektronegative Atome wie z.B. Sauerstoff, Stickstoff, Schwefel oder auch Halogene. Die Bindungsenergien für Wasserstoffbrücken liegen zwischen 12 und 38 kJ/mol [114]. Bei intermolekularen Wasserstoffbrücken zwischen zwei Komplexpartnern liefern sie somit einen signifikanten Beitrag zur Gesamtbindungsenergie des Komplexes. Gleichzeitig ist die Energiebarriere zur Überwindung einer Wasserstoffbrückenbindung noch hinreichend gering, um eine schnelle Dissoziation der Komplexpartner zu ermöglichen. Zudem tragen Wasserstoffbrücken zur Spezifität der Komplexbildung bei, da Donatoren und Akzeptoren in der Kontaktregion nur bei optimaler Komplementarität Wasserstoffbrücken zum Komplexpartner ausbilden können. Nicht abgesättigte Donatoren oder Akzeptoren in der Binderegion eines Komplexes tragen erheblich zu dessen Destabilisierung bei, da sie durch Dissoziation des Komplexes Wasserstoffbrücken mit dem Lösungsmittel ausbilden können.

Ist die genaue Struktur des Komplexes aufgeklärt, so können Wasserstoffbrücken zwischen den Komplexpartnern explizit untersucht werden. Ist allerdings nur die Struktur der einzelnen Komplexpartner zugänglich, so kann die Untersuchung der dem Lösungsmittel zugewandten Atomgruppen hinsichtlich ihrer Wasserstoffbrückenbildungsfähigkeit Aufschluß über Stellen mit potentiellen Wasserstoffbrücken geben. Zur Darstellung dieser Wasserstoffbrückenbindungsfähigkeit auf der molekularen Oberfläche werden im folgenden zwei Verfahren vorgestellt.

2.5.3.1 Wasserstoffbrückenbildungsfähigkeit

Von Heiden [81] wurde ein Verfahren entwickelt, mit dem die Fähigkeit einzelner Atome zur Bildung von Wasserstoffbrücken auf der Moleküloberfläche dargestellt werden kann. Dabei wird zunächst jedem Atom einer der folgenden Werte zugewiesen, der angibt ob es sich um einen Wasserstoffakzeptor oder -donor handelt:

-1 = potentieller Wasserstoffakzeptor

0 = keine Wasserstoffbrückenbildung

+1 = potentieller Wasserstoffdonor / korrespondierendes Wasserstoffatom

Anschließend wird jedem Oberflächenpunkt das ihm nächstgelegene Atom (kürzester Abstand zur van-der-Waals-Oberfläche des Atoms) ermittelt und dessen vorher zugewiesener Zahlenwert farbcodiert auf der Oberfläche abgebildet. Die dargestellte Methode eignet sich zur qualitativen Beurteilung, an welchen Stellen des Moleküls potentielle Wasserstoffbrücken mit dem umgebenden Solvens bzw. dem Komplexpartner ausgebildet werden können.

2.5.3.2 Wasserstoffakzeptoren- und Wasserstoffdonatordichte

Keil [35] erweiterte den Ansatz von Heiden, indem er nicht nur das nächstgelegene Atom zur Bestimmung der Wasserstoffbrückenbildungsfähigkeit benutzt, sondern auch die benachbarte Umgebung der Oberflächenpunkte bei der Berechnung berücksichtigt. Dazu wird wie unter Abschnitt 2.5.3.1 beschrieben zunächst allen Oberflächenpunkten die Wasserstoffbrückenbindungsfähigkeit ihres nächstgelegenen Atoms zugewiesen. Oberflächenpunkte, die demselben Atom zugeordnet sind, werden zu einer Teiloberfläche (*Patch*) zusammengefaßt, die dann einen Donatoren- oder Akzeptoren-*Patch* darstellt. Für jeden Punkt wird anschließend die Anzahl der Donatoren- bzw. Akzeptoren-*Patches* bestimmt, die sich vollständig innerhalb eines bestimmten *Cutoff*-Abstandes von dem Oberflächenpunkt befinden. Für *Patches*, die nicht vollständig innerhalb des *Cutoff*-Abstandes liegen, wird nur der Anteil berücksichtigt, der sich innerhalb des *Cutoffs* befindet. Der Quotient aus der Summe der Anteile aller Donatoren- bzw. Akzeptoren-*Patches* $n_{D/A}$ und der vom *Cutoff*-Radius eingeschlossenen Moleküloberfläche a ergibt die Donatoren- bzw. Akzeptordichte $r_{D/A}$ des Oberflächenpunktes i :

$$\mathbf{r}_{D/A}(i) = \frac{\sum_k n_{D/A}(k)}{a(i)} . \quad (2.7)$$

Somit läßt sich eine Dichte für Wasserstoffdonatoren, -akzeptoren oder auch für beide gleichzeitig bestimmen. Durch Variation des *Cutoff*-Abstandes läßt sich festlegen, wie stark weiter entfernte Atome bei der Berechnung der Dichte miteinbezogen werden. In dieser Arbeit wurde ein *Cutoff* von 2.5 Å verwendet, um die lokale Umgebung stärker zu berücksichtigen.

2.5.4 Salzbrücken

Die elektrostatische Bindung zweier entgegengesetzt geladener Gruppen innerhalb eines Proteins oder zwischen zwei Komplexpartnern nennt man Ionenpaarbindung oder Salzbrücke. Die beiden geladenen Gruppen sind bei optimaler Ausrichtung ca. 2.8 Å voneinander entfernt. Häufig sind bei der Ausbildung von Salzbrücken auch Wasserstoffbrücken beteiligt. Wie die Wasserstoffbrückenbindung tragen auch Salzbrücken erheblich zur Spezifität der Komplexbildung bei. Eine isolierte Ladung in der Binderegion, die keine Salzbrücke zum Komplexpartner ausbilden kann, führt zur Destabilisierung des Komplexes. Nussinov *et al.* [26] konnten zeigen, daß bei Protein-Protein-Komplexen in der Binderegion im Durchschnitt zwei Salzbrücken ausgebildet werden. Die Untersuchung von Salzbrücken stellt somit ein weiteres wichtiges Kriterium bei der Beschreibung von Proteinkomplexen dar. In dieser Arbeit wurde daher ein Ansatz entwickelt, potentielle Salzbrückenbildner auf der molekularen Oberfläche der separierten Komplexpartner darzustellen. Die Methode ähnelt dem unter Abschnitt 2.5.3.1 vorgestellten Verfahren zur Darstellung der Wasserstoffbrückenbildungsfähigkeit. Zunächst wird anhand der in den Strukturdaten vorgegebenen Protonierung der potentiell geladenen Aminosäuren Arginin, Histidin, Lysin, Glutamin- und Asparaginsäure sowie des N- und C-Terminus bestimmt, ob diese geladen vorliegen. Soll die in den Strukturdaten enthaltene Protonierung ignoriert werden, oder liegt keine Information über die Protonierung vor, besteht die Möglichkeit, alle oben genannten Gruppen als geladen zu markieren. Dies ermöglicht die schnelle Vorhersage *potentieller* Salzbrückenbildner, ohne aufwendige Berechnung von pK_s -Werten

zur Vorhersage des tatsächlichen Protonierungszustandes durchführen zu müssen. Anschließend wird jedem Atom einer der folgenden Werte zugewiesen:

-1 = negativ geladene Salzbrückenbildner

Asparagin-/Glutaminsäure: Sauerstoffe der Carboxygruppe der Seitenkette

C-Terminus: Sauerstoffe der Carboxygruppe der Hauptkette

0 = ungeladene Atome bzw. Heteroatome unbekannter Gruppen

+1 = positiv geladene Salzbrückenbildner

Arginin: Stickstoffe und daran gebundene Wasserstoffe der Guanidinogruppe

Histidin: Stickstoffe und daran gebundene Wasserstoffe der Imidazolylgruppe

Lysin: Stickstoff und Wasserstoffe der Aminogruppe der Seitenkette

N-Terminus: Stickstoff und Wasserstoffe der Aminogruppe der Hauptkette

Für die Darstellung der Werte auf der Oberfläche wird jedem Oberflächenpunkt das ihm nächstgelegene Atom (kürzester Abstand zur van-der-Waals-Oberfläche des Atoms) ermittelt und dessen Zahlenwert farbcodiert auf der Oberfläche wiedergegeben. Somit kann qualitativ untersucht werden, an welchen Stellen des Moleküls die Ausbildung von Salzbrücken möglich ist.

2.5.5 Topographische Analyse der Oberfläche

Die Ausbildung der oben besprochenen elektrostatischen und hydrophoben Wechselwirkungen zwischen zwei Komplexpartnern hängt stark von deren sterischer Komplementarität in der Binderegion ab. Nur wenn die beiden Komplexpartner wie Schlüssel und Schloß ineinander passen, kann der Komplex durch intermolekulare Wechselwirkungen optimal stabilisiert werden. Es ist daher hilfreich die Topographie der Binderegion eines Moleküls beschreiben zu können, um beispielsweise konvexe von konkaven Bereichen zu unterscheiden. Die in dieser Arbeit verwendeten Ansätze zur Berechnung der Molekültopographie mit Hilfe molekularer Oberflächen werden im folgenden kurz erläutert.

2.5.5.1 Oberflächenkrümmung, Shapeindex und Curvedness

Die lokale Topographie einer molekularen Oberfläche läßt sich für jeden Oberflächenpunkt durch die beiden sogenannten kanonischen Krümmungen quantitativ beschreiben [115]. Sie sind als die Eigenwerte der lokalen Hesse-Matrix definiert und spiegeln eine atomare Auflösung der Topographie wider. Bei der Betrachtung von Protein-Protein-Komplexen ist allerdings die Komplementarität größerer, zusammenhängender Molekülbereiche von Bedeutung. Von Zachmann *et al.* [116] wurde daher ein Ansatz vorgeschlagen, die globalen Krümmungen eines zusammenhängenden Oberflächenbereichs vordefinierter Größe zu bestimmen. Exner [38] erweiterte den Ansatz und entwickelte ein Verfahren, bei dem die globalen Krümmungen für einen Oberflächenpunkt unter Berücksichtigung aller benachbarten Oberflächenpunkte innerhalb eines bestimmten Abstandes (*Cutoff*-Radius) berechnet werden können. Dabei wird zunächst ein Paraboloid numerisch an die ausgewählte Punktmenge angepaßt. Die Hauptkrümmungen des Paraboloids definieren die globalen Krümmungen, die dem jeweiligen Oberflächenpunkt zugeordnet werden. Die Krümmungen können anschließend zu zwei charakteristischen topographischen Indizes zusammengefaßt werden, von denen einer die Form (konvex / konkav) und einer die Neigung (Stärke der Krümmung) der Oberfläche beschreibt.

Zur Beschreibung der Oberflächenform eignet sich der von Duncan & Olson [117,118] vorgeschlagene Ansatz des *Shapeindex* (s), der auf die oben beschriebenen globalen Krümmungen (k_1, k_2) angewendet werden kann:

$$s = -\frac{p}{2} \cdot \arctan\left(\frac{k_1 + k_2}{k_1 - k_2}\right). \quad (2.8)$$

Die Werte des *Shapeindex* liegen zwischen $s = -1$ für konkave Bereiche ($k_1 = k_2 > 0$) und $s = 1$ für konvexe Bereiche ($k_1 = k_2 < 0$). Zur Verdeutlichung der Bedeutung des *Shapeindex* kann die folgende grobe Zuordnung zu geometrischen Formen dienen:

$$\text{Loch } (s = -1) \rightarrow \text{Spalt } (s = -0.5) \rightarrow \text{Sattel } (s = 0) \rightarrow \text{Grat } (s = 0.5) \rightarrow \text{Pfropf } (s = 1)$$

Um Ungenauigkeiten bei der Berechnung auszugleichen, schlägt Exner [38] vor, den *Shapeindex* bei betragsmäßig kleinen Werten von k_1 und k_2 durch einen Zusatzterm zu

korrigieren sowie betragsmäßig große Werte von k_1 und k_2 auf einen Maximal- (0.8 \AA^{-1}) bzw. Minimalwert (-0.8 \AA^{-1}) anzupassen. Diese Modifikationen wurden auch in der vorliegenden Arbeit verwendet. Für Einzelheiten wird auf die Publikation von Exner [38] verwiesen.

Die Oberflächenneigung lässt sich durch die von Duncan & Olson [117,118] eingeführte *Curvedness* (R) beschreiben:

$$R = \sqrt{\frac{1}{2} \cdot (k_1^2 + k_2^2)}. \quad (2.9)$$

Je größer der Wert für R , umso stärker ist die Oberflächenneigung. Da der Wertebereich der globalen Krümmungen in dem Ansatz von Exner beschränkt wurde, beträgt der Maximalwert der *Curvedness* $R = 0.8 \text{ \AA}^{-1}$. Die Beschreibung der Topographie einer Oberfläche kann nun erfolgen, indem jedem Oberflächenpunkt ein Wert für dessen *Shapeindex* bzw. *Curvedness* zugewiesen wird.

2.5.5.2 Tiefeninformation

Neben der Bestimmung von konvexen und konkaven Proteinregionen ist die Kenntnis über tiefe Spalten oder Taschen in der Binderegion zur Beschreibung von Proteinkomplexen hilfreich. Tiefe Einbuchtungen stellen häufig die Bindestelle für kleinere Liganden dar oder bilden das aktive Zentrum des Proteins. Im Fall von Protein-Protein-Komplexen können Einbuchtungen in der Binderegion Wassermoleküle aufnehmen und diese so vom restlichen Solvens separieren. Zum Auffinden solcher Einbuchtungen existieren mehrere geometrische Ansätze, die mit Hilfe von Probekugeln die Proteinkoordinaten absuchen [53,55]. Das in dieser Arbeit verwendete Verfahren von Keil [35] beruht auf der Berechnung zweier molekularer Oberflächen mit unterschiedlich großen Abrollradien. Die Referenzoberfläche (1) besitzt einen großen Abrollradius ($r_1 = 6.0 \text{ \AA}$) und bedeckt alle Taschen und Kanaleingänge, während die zweite Oberfläche (2) die übliche molekulare Oberfläche ($r_2 = 1.4 \text{ \AA}$) darstellt. Jedem Oberflächenpunkt der molekularen Oberfläche (2) wird der kürzeste Abstand zur Referenzoberfläche (1) zugeordnet, der die „Tiefe“ der Tasche angibt. Bei diesem Ansatz werden Oberflächenbereiche als Tasche definiert, wenn der Abstand zur Referenzoberfläche (1) größer als 2.5 \AA ist.

3 Vorhersage von Wasserpositionen in Proteinkomplexen

In diesem Kapitel wird ein neues Verfahren vorgestellt, das - ausgehend von der Kristallstruktur eines Protein-Protein-Komplexes - Kavitäten in der Binderegion des Komplexes identifiziert, die groß genug sind, um mit Wassermolekülen besetzt zu werden. Für die Kavitäten können darüber hinaus die Anzahl der Wassermoleküle, die in ihnen Platz finden, sowie die potentiellen Positionen dieser Wassermoleküle vorhergesagt werden. Damit der Algorithmus optimiert und getestet werden kann, ist ein Datensatz von Kristallstrukturen unterschiedlicher Protein-Protein-Komplexe notwendig. Die Zusammenstellung und Aufbereitung passender Kristalldaten aus der *Protein Data Bank* [3] sowie die Optimierung und Validierung des neuen Verfahrens mit Hilfe experimenteller und modellierter Daten von Wassermolekülen wird im folgenden ausführlich besprochen.

3.1 Auswahl des Datensatzes

Voraussetzung für die neue Methode zur Identifikation potentieller Wasserpositionen in der Binderegion von Protein-Protein-Komplexen ist die Kenntnis über die dreidimensionale Struktur des Komplexes. Mit Hilfe von Röntgenbeugungsuntersuchungen an Kristallen oder unter Verwendung der 2D-Kernspinresonanzspektroskopie können die Koordinaten der Atomschwerpunkte experimentell bestimmt werden. Viele der so bestimmten Strukturen von Proteinen, Proteinkomplexen und Nukleinsäuren sind in der *Protein Data Bank* (PDB) [3] gespeichert und über das Internet frei zugänglich. Die in dieser Arbeit verwendeten Kristallstrukturen entstammen alle der PDB und sind durch ihren vierstelligen, PDB-eigenen Identifikationscode (PDB-ID) gekennzeichnet.

Gegenstand der Untersuchungen ist die Analyse der Binderegion von Protein-Protein-Komplexen. Aus diesem Grund werden Proteinkomplexe mit kleinen oder nicht-peptidischen Liganden bei der Zusammenstellung des Datensatzes nicht berücksichtigt. Das hier beschriebene Verfahren zur Identifikation potentieller Wasserpositionen ist allerdings prinzipiell auch auf solche Proteinkomplexe anwendbar. Als Protein-Protein-Komplexe werden ausschließlich Homodimere, Antigen-Antikörper- oder Enzym-Substrat/Inhibitor-Komplexe angesehen. Besteht ein Komplexpartner aus mehreren Proteinketten, werden diese als eine funktionelle Einheit angesehen und zusammengefaßt.

In einigen PDB-Einträgen liegen mehrere Strukturen des gleichen Komplexes vor. In solchen Fällen wird jede der vorliegenden Komplexstrukturen als eigenständiger Komplex in den Datensatz integriert. Der verwendete Datensatz besteht aus insgesamt 111 Protein-Protein-Komplexen, die aus 91 PDB-Einträgen stammen. Darunter befinden sich 32 Antigen-Antikörper-Komplexe, 70 Enzymkomplexe sowie 9 Dimere. Die Größe der einzelnen Komplexpartner reicht von 10 Aminosäuren im Fall eines Erkennungs-Peptids im Komplex mit dem Anti-Prionen F_{ab}-Fragment 3F4 (1cu4) bis zu 828 Aminosäuren im Fall von TAQ-DNA-Polymerase im Komplex mit dem F_{ab}-Fragment TP7 (1bgx). Eine Aufstellung der PDB-Einträge des Datensatzes ist in Tabelle 3.1 zu finden. Jeder Eintrag des Datensatzes enthält neben den Proteinstrukturen auch Koordinaten von Kristallwassermolekülen, so daß später ein Vergleich von theoretisch ermittelten und experimentellen Wasserdaten möglich ist. Von den verwendeten Strukturen weist jeder Komplexpartner eine Gesamtsequenzlänge von mindestens 10 Aminosäuren auf. Die Binderegion jedes Komplexpartners besteht ausschließlich aus den 20 Standard-aminosäuren und beinhaltet weder Metallatome noch nicht-peptidische Liganden oder Aminosäurederivate (sog. *Heteroatome*). Dies erleichtert die spätere Bearbeitung und Analyse der Binderegion, da sichergestellt ist, daß in dem verwendeten *CHARMM*-Kraftfeld [103] für alle Atome der Binderegion Kraftfeldparameter existieren. Eine Ausnahme stellt der PDB-Eintrag 1stf dar, der einen carboxymethylierten Cysteinrest in der Binderegion besitzt, für den aus einer vorangegangenen Arbeit Parameter in das Kraftfeld integriert werden konnten [80].

Enzyme, die ohne Substrat oder Inhibitor vorliegen, aber aus mehreren Proteinketten bestehen, stellen prinzipiell natürlich ebenfalls Proteinkomplexe dar. Das gleiche gilt für Antikörperfragmente, die auch ohne Antigen einen Komplex darstellen, da sie aus einer schweren und einer leichten Proteinkette bestehen. Diese Art von Komplexen werden in der vorliegenden Arbeit jedoch nicht als Protein-Protein-Komplex berücksichtigt.

Dimere: 9								
PDB ID	Ketten ID		PDB ID	Ketten ID		PDB ID	Ketten ID	
1cdt	A	B	1ypi	A	B	2tsc	A	B
1msb	A	B	2ccy	A	B	3sdp	A	B
1pp2	R	L	2rus	A	B	5rub	A	B
Antikörper: 32								
PDB ID	Ketten ID		PDB ID	Ketten ID		PDB ID	Ketten ID	
1a2y	A/B	C	1igc	H/L	A	1nmc_1	N	H/L
1arl	A/B	C/D	1jrh	H/L	I	1nmc_2	A	B/C
1bgx	T	H/L	1kb5	A/B	H/L	1oak	H/L	A
1bog	A/B	C	1kip	A/B	C	1osp	H/L	O
1bql	H/L	Y	1kiq	A/B	C	1qfu	A/B	H/L
1cic	A/B	C/D	1kir	A/B	C	1vfb	A/B	C
1cu4	H/L	P	1mlc_1	A/B	E	2iff	H/L	Y
1dqj	A/B	C	1mlc_2	C/D	F	2jel	H/L	P
1ejo	H/L	P	1nca	N	H/L	3hfl	H/L	Y
1eo8	A/B	H/L	1ncb	N	H/L	3hfm	H/L	Y
1fns	H/L	A	1ncd	N	H/L			
Enzyme: 70								
PDB ID	Ketten ID		PDB ID	Ketten ID		PDB ID	Ketten ID	
1acb	E	I	1ca0_1	A/B/C	D	1tab	E	I
1avw	A	B	1ca0_2	F/G/H	I	1taw	A	B
1avx	A	B	1cbw_1	A/B/C	D	1tec	E	I
1b27_1	A	D	1cbw_2	F/G/H	I	1tgs	Z	I
1b27_2	B	E	1cgi	E	I	1tpa	E	I
1b27_3	C	F	1cgj	E	I	1udi	E	I
1b2s_1	A	D	1cho	E	I	2btc	E	I
1b2s_2	B	E	1cse	E	I	2kai	A/B	I
1b2s_3	C	F	1eai_1	A	C	2pcb	A	B
1b2u_1	A	D	1eai_2	B	D	2pcc_1	A	B
1b2u_2	B	E	1f2s	E	I	2pcc_2	C	D
1b2u_3	C	F	1fin_1	A	B	2ptc	E	I
1b3s_1	A	D	1fin_2	C	D	2sec	E	I
1b3s_2	B	E	1fle	E	I	2sic	E	I
1b3s_3	C	F	1fss	A	B	2sni	E	I
1bgs_1	A	E	1hia_1	A/B	I	2sta	E	I
1bgs_2	B	F	1hia_2	X/Y	J	2stb	E	I
1bgs_3	C	G	1mct	A	I	2tgp	Z	I
1brb	E	I	1ppe	E	I	3sgb	E	I
1brc	E	I	1ppf	E	I	3tgi	E	I
1brs_1	A	D	1sbn	E	I	3tgj	E	I
1brs_2	B	E	1sbw	A	I	4sgb	E	I
1brs_3	C	F	1sib	E	I			
1bzx	E	I	1stf	E	I			

Tab. 3.1: PDB-Einträge des Datensatzes von Protein-Protein-Komplexen. Einträge mit mehreren Strukturen sind durch einen Unterstrich markiert. Die Namen der Proteinketten, aus denen die einzelnen Komplexpartner bestehen, sind ebenfalls aufgeführt.

3.2 Bearbeitung der Kristallstrukturen

Die aus der PDB entnommenen Kristallstrukturen enthalten in den meisten Fällen keine Wasserstoffatome. Darüber hinaus fehlen in manchen Strukturen die Koordinaten einzelner Schweratome. Außerdem kann es vorkommen, daß für Atome mehrere alternative Koordinaten mit unterschiedlichen Besetzungswahrscheinlichkeiten angegeben werden. Aus diesen Gründen ist es notwendig, die Daten aus der PDB zuerst zu bearbeiten, um ein korrektes Strukturmodell aufbauen zu können. Dies geschieht für jeden PDB-Eintrag des Datensatzes mit einem von Keil [35,77] entwickelten, automatisierten Verfahren. Dabei wird zunächst der Originaleintrag aus der PDB mit dem Programm *blackboxing* analysiert. Dieses Programm identifiziert alle in der PDB enthaltenen Proteinketten, Liganden und Wassereinträge und erstellt ein Programmskript (*blackbox.sh*), mit dem alle folgenden Schritte automatisiert durchgeführt werden. Besteht ein Komplexpartner aus mehreren Proteinketten, muß diese Information allerdings manuell in das Skript integriert werden, da ansonsten alle Proteinketten als separate Einheiten behandelt werden. So ist es beispielsweise sinnvoll, die schwere und leichte Kette eines Antikörpers am Ende des Skriptdurchlaufes zu einer einzelnen Datei zu vereinigen, so daß der komplette Antikörper einen Komplexpartner darstellt. Welche Untereinheiten zu einem Komplexpartner zusammengefaßt werden, ist in Tabelle 3.1 aufgeführt. Die Berechnung fehlender Schweratome und die Addition von Wasserstoffatomen geschieht mit dem Programm *CHARMM* in der Version 24g2 [103]. Alle notwendigen Datenkonvertierungen sowie die Erstellung der *CHARMM*-Eingabedatei erfolgen automatisch durch das *blackbox.sh*-Skript. Für den Protonierungszustand saurer und basischer Aminosäuren werden die folgenden Annahmen gemacht:

- *Arginin* und *Lysin* sind einfach positiv geladen
- *Asparagin*- und *Glutaminsäure* sind einfach negativ geladen
- *Histidin* ist neutral; bei der Bindung an Metallionen wird das an der Bindung unbeteiligte Stickstoffatom protoniert
- *Cysteine*, die laut Abstandsberechnungen Disulfidbrücken ausbilden, werden miteinander verbunden und bleiben unprotoniert

Die Energieminimierung der fehlenden Schweratome und Wasserstoffatome erfolgt mit dem *Adopted-basis-set-Newton-Raphson*-Verfahren aus *CHARMM* in maximal 2000 Schritten oder bis die Änderung der Gesamtenergie einen Wert von 10^{-8} kcal/mol unterschreitet. Die in der Kristallstruktur enthaltenen Atompositionen bleiben bei der Energieminimierung auf ihren vorgegebenen Positionen fixiert. Am Ende des Skriptes werden die beiden Komplexpartner sowie die experimentell gefundenen Kristallwassermoleküle in je einer Datei im PDB-Format abgelegt. Somit wird jeder Komplex des Datensatzes durch jeweils drei Dateien beschrieben, welche die Grundlagen für alle weiteren Untersuchungen bilden.

3.3 Beschreibung des Verfahrens zur Vorhersage von Wasserpositionen

Die Berechnung potentieller Wasserpositionen in der Binderegion von Protein-Protein-Komplexen beruht auf einem rein geometrischen Ansatz. Zunächst wird für jeden einzelnen Komplexpartner die Connolly-Oberfläche mit dem Programm *fume* [35] bestimmt. Der Radius der Probekugel beträgt hierbei 1.3 Å („innerer Radius“). Dieser Wert ist für die Problemstellung optimiert und wird unter Abschnitt 3.4 näher erläutert. Für den Gesamtkomplex wird ebenfalls eine Connolly-Oberfläche erzeugt, bei der für die Probekugel allerdings ein größerer Radius von 1.8 Å („äußerer Radius“, Abschnitt 3.4.3) verwendet wird (Abb. 3.1-A). Mit dem größeren äußeren Radius wird erreicht, daß kleine Kanäle, die eine Verbindung zwischen Kavitäten im Innern der Binderegion und der umgebenden Wasserphase (*Bulk*-Phase) herstellen, durch die berechnete Komplexoberfläche geschlossen werden. Die Größe der beiden Radien kann je nach Fragestellung frei gewählt werden. Der Einfluß unterschiedlicher Werte für den inneren und äußeren Radius auf das Verfahren wird in Abschnitt 3.4 diskutiert. Bei der Oberflächengenerierung kann es vorkommen, daß sich innerhalb der das Protein umschließenden Oberfläche Hohlräume befinden, die wiederum von einer Oberfläche umgeben sind. Mit dem Programm *splitsld* (Keil, persönliche Mitteilung) wird sichergestellt, daß diese kleinen Oberflächen innerhalb der großen molekularen Oberfläche entfernt werden. Die so bereinigten molekularen Oberflächen der beiden Komplexpartner und des Gesamtkomplexes dienen neben den Atomkoordinaten als Eingabe für den Algorithmus zur Berechnung potentieller Wassermoleküle, der in dem Programm *space* umgesetzt wird.

Der Algorithmus ist ein gitterbasiertes Verfahren, bei dem zunächst der Gesamtkomplex in ein reguläres, dreidimensionales Gitter mit einem Punktabstand von 0.3 \AA gesetzt wird (Abb. 3.1-B). Anschließend werden alle Punkte, die sich innerhalb der Oberflächen der Komplexpartner und außerhalb der Gesamtkomplexoberfläche befinden verworfen (Abb. 3.1-C). Punkte, die sich nicht in der Binderegion befinden werden ebenfalls gelöscht. Die Binderegion wird durch diejenigen Atome definiert, die nicht weiter als 8 \AA von ihrem Komplexpartner entfernt sind. Die resultierenden Punkte stellen die Hohlräume nahe der Binderegion des Proteinkomplexes dar. Um Hohlräume zu identifizieren, die groß genug sind, mindestens ein Wassermolekül aufzunehmen, wird auf jeden Punkt eine Probekugel mit Radius 1.3 \AA platziert. Diese Probekugel („Wasserkugel“) stellt für das vorliegende Problem in guter Näherung die Ausdehnung eines Wassermoleküls dar. Die Optimierung des Probekugelradius wird in Abschnitt 3.4 diskutiert. Alle Punkte, deren Abstand zum nächsten Atom der Binderegion kleiner als die Summe des van-der-Waals-Radius dieses Atoms und der Probekugel sind, werden gelöscht. Im Falle von Wasserstoffbrückenbindungen können sich Donor und Akzeptor näher kommen, als mit dem Hartkugelmodell möglich ist. Um diesen verkürzten Abständen Rechnung zu tragen, sind bei den folgenden Ausnahmen auch kürzere Abstände zwischen den Punkten und ihren nächsten Atomen erlaubt: Wasserstoff 1.6 \AA , Stickstoff 2.6 \AA , Sauerstoff 2.6 \AA . Alle resultierenden Punkte stellen nun Positionen dar, auf denen ein Wassermolekül platziert werden kann (Abb. 3.1-D). Anschließend werden die Punkte durch einen hierarchischen, agglomerativen Clusteralgorithmus in räumlich voneinander getrennte Cluster eingeteilt, von denen jeder eine abgeschlossene Kavität darstellt. Durch Verwendung des *Single-Linkage*-Verfahrens werden Cluster gebildet, in denen jeder Gitterpunkt eines Clusters nicht weiter als 2.6 \AA (doppelter Probekugelradius) von mindestens einem anderen, im selben Cluster befindlichen Gitterpunkt, entfernt ist. Cluster, deren kürzester Abstand zu einem der Komplexpartner größer als die Summe aus Gitterabstand, Probekugelradius und van-der-Waals-Radius des nächstgelegenen Atoms ist, liegen nicht zwischen den Komplexpartnern, sondern am Rand der Binderegion und werden verworfen. Die übrigen Cluster repräsentieren räumlich voneinander getrennte Hohlräume in der Binderegion, in denen jeder Punkt eine potentielle Wasserposition darstellt (Abb. 3.1-D).

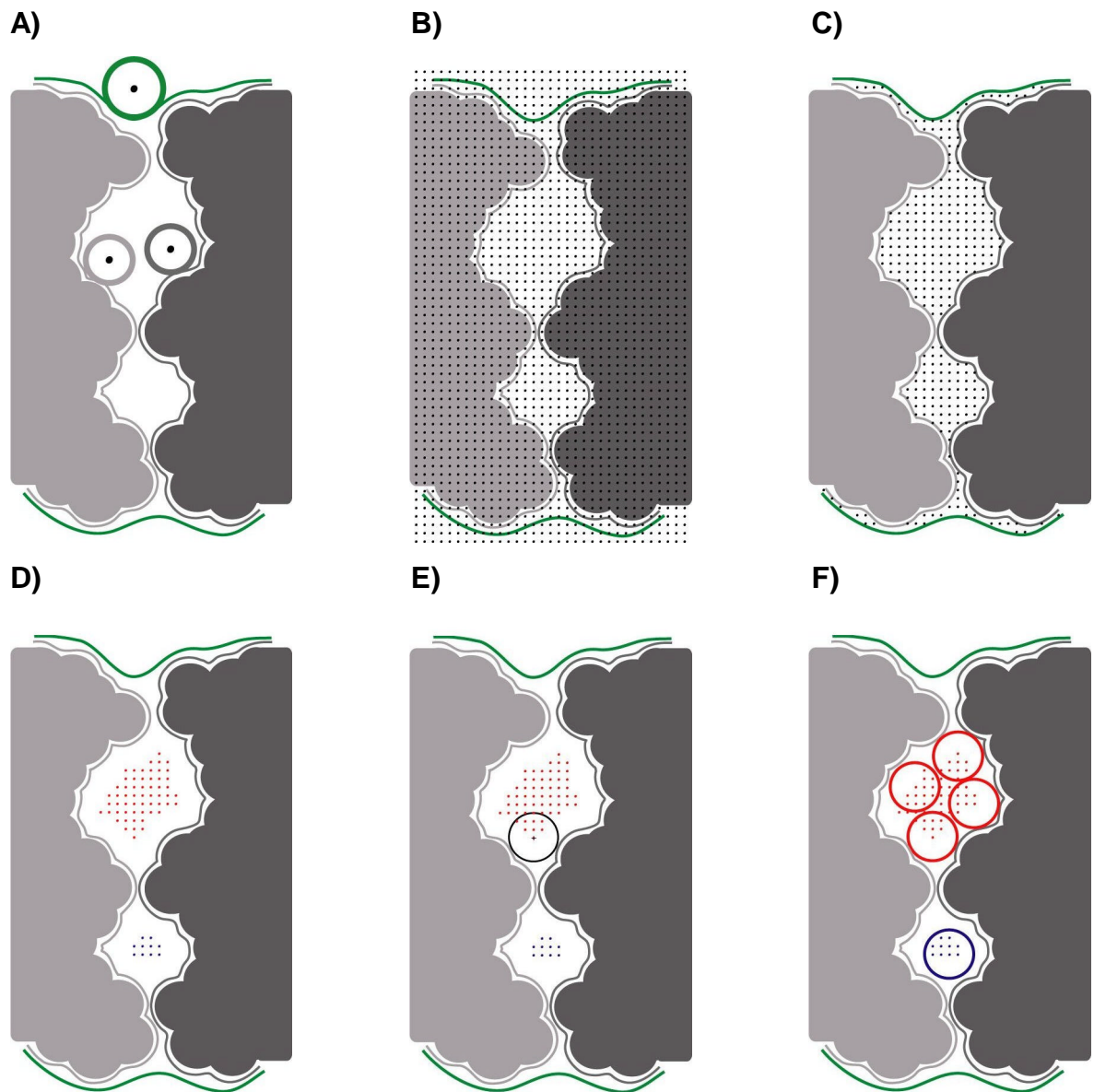


Abb. 3.1: Schematischer Verlauf des Verfahrens zur Vorhersage von Wassermolekülen in der Binderegion von Protein-Protein-Komplexen. **A)** Die Connolly-Oberflächen der einzelnen Komplexpartner (grau, Abrollradius 1.3 Å) und des Gesamtkomplexes (grün, Abrollradius 1.8 Å) werden erzeugt. **B)** Der Komplex wird in ein reguläres, dreidimensionales Gitter gelegt (Gitterabstand 0.3 Å). **C)** Punkte innerhalb der Oberflächen der Komplexpartner und außerhalb der Oberfläche des Gesamtkomplexes werden verworfen. **D)** Punkte, auf denen eine Kugel vom Radius 1.3 Å ohne Überlappung mit umgebenden Atomen platziert werden kann, stellen potentielle Wasserpositionen dar; räumlich benachbarte Punkte werden zu einem Cluster zusammengefaßt (rot bzw. blau). **E)** Die erste „Wasserkugel“ zum Auffüllen eines Clusters wird auf den am weitesten vom Schwerpunkt entfernten Punkt platziert. **F)** Die Anzahl der in den Clustern untergebrachten Kugeln entspricht der vorhergesagten Wasseranzahl in den Kavitäten der Binderegion

Um die Anzahl der Wassermoleküle vorherzusagen, die von den Clustern aufgenommen werden können, wird jeder Cluster mit Probekugeln vom Radius 1.3 \AA („Wasserkugeln“) aufgefüllt, die Wassermoleküle repräsentieren. Die erste Kugel wird auf den vom Clusterschwerpunkt am weitesten entfernten Punkt platziert (Abb. 3.1-E). Anschließend wird der nächstmögliche Punkt des Clusters gesucht, auf den eine zweite Kugel gesetzt werden kann, ohne mit der ersten Kugel zu überlappen. Dieser Vorgang wird sukzessive durchgeführt, bis keine weitere Kugel mehr in dem Cluster Platz findet. Die Anzahl der geometrisch gesetzten Kugeln entspricht der potentiellen Wasseranzahl in der Kavität (Abb. 3.1-F). Die durch das beschriebene Verfahren berechneten Kugelpositionen werden als Datei im PDB-Format gespeichert und können nun als Startpunkt für Sauerstoffatome beim Aufbau von Wassermolekülen dienen. Durch Addition von Wasserstoffatomen und anschließender Energieminimierung der Wassermoleküle können die Wasserpositionen noch präzisiert werden.

3.4 Parametrisierung des Verfahrens

Anzahl und Positionen der Wassermoleküle, die mit dem oben besprochenen Verfahren berechnet werden, hängen von den verwendeten Radien bei der Oberflächengenerierung sowie dem Radius der Probekugel ab, die ein Wassermolekül simuliert. Der „äußere“ Radius bei der Oberflächengenerierung des Gesamtkomplexes legt fest, ob Bereiche, die mit dem *Bulk*-Wasser durch einen Kanal in Verbindung stehen, noch als „abgeschlossene“ Kavität definiert werden, oder ob sie als dem *Bulk*-Wasser zugewandte Oberflächenbereiche angesehen werden. Dabei ist der optimale Radius anwendungsspezifisch, so daß der Anwender je nach Zielsetzung und untersuchtem System entscheiden kann, ab welcher Größe der Zugangsöffnung die Bereiche noch als Kavität behandelt werden sollen. Anders verhält es sich mit dem „inneren“ Radius bei der Generierung der einzelnen Oberflächen der Komplexpartner und dem Radius der Probekugel beim Auffüllen der Kavitäten. Diese beiden Radien werden bei dem Verfahren gleich groß gewählt, da sie beide die Größe eines Wassermoleküls beschreiben. Im folgenden wird daher nur noch von einem „Wasserradius“ gesprochen, der beide Radien beschreibt. Die Größe dieses Wasserradius ist bestimmend dafür, welche Kavitäten gefunden werden und wie viele Wassermoleküle darin Platz finden. Durch Variation des Wasserradius kann erreicht werden, daß die Geometrie eines Wassermoleküls in einer Kavität möglichst genau wiedergegeben wird.

Zur Optimierung dieses Radius ist ein Datensatz an Kavitäten notwendig, von denen die Anzahl der darin befindlichen Wassermoleküle bekannt ist. Da die experimentell ermittelten Kristallwasserdaten aus den Komplexstrukturen des Datensatzes aus Tab. 3.1 die Wasserbesetzung der Kavitäten nur unvollständig beschreiben, wird für die Optimierung des Radius ein reduzierter Testdatensatz verwendet, bei dem die optimale Wasserbesetzung von Kavitäten mit Hilfe von Molekulardynamiksimulationen (MD-Simulationen) bestimmt wurde (P. Monecke, persönliche Mitteilung). Im folgenden wird kurz die Vorgehensweise zur Berechnung des reduzierten Datensatzes skizziert. Anschließend wird die Optimierung des Wasserradius besprochen und die Auswirkung unterschiedlicher äußerer Radien auf den Algorithmus diskutiert.

3.4.1 Erstellung eines Datensatzes an Kavitäten mit optimaler Wasseranzahl

Die Berechnung der optimalen Wasseranzahl in Kavitäten für einen Datensatz an Proteinkomplexen mit Hilfe von MD-Simulationen wurde von P. Monecke (persönliche Mitteilung) durchgeführt. Der reduzierte Datensatz ist eine Teilmenge des unter Abschnitt 3.1 vorgestellten Datensatzes. Er besteht aus 26 Antikörper-Komplexen, die insgesamt 211 Kavitäten enthalten, denen ihre optimale Anzahl an Wassermolekülen zugeordnet ist (Tab. 3.2). Zur Berechnung dieser Wasseranzahl werden zunächst wie unter 3.2 beschrieben die Kristallstrukturen der Komplexe bearbeitet und anschließend die Kavitäten in der Binderegion mit dem unter 3.3 dargestellten Verfahren identifiziert. Als „äußerer“ Radius werden 1.8 \AA , als „innerer“ Radius 1.2 \AA gewählt. Zum Auffüllen der Kavitäten mit Wassermolekülen wird allerdings ein Probekugelradius von 0.8 \AA verwendet, so daß die Kavitäten mit einer größeren Wasseranzahl besetzt werden können, als real tatsächlich möglich wäre. Dies dient dazu, auch Systeme mit sehr hoher Wasserbesetzung simulieren zu können. Damit die Wassermoleküle die Kavität während der MD-Simulation nicht verlassen können, wird jeder potentielle Kanal ins Proteininnere oder ins *Bulk*-Wasser mit sogenannten „Dummy“-Atomen verschlossen, die neutral sind und keine anziehenden Wechselwirkungsenergien aufweisen. Da jede Kavität einzeln simuliert wird, werden alle Proteinatome, die weiter als 10 \AA von einem Wassermolekül der Kavität entfernt sind, für die Simulation gelöscht. Die MD-Simulation jeder Kavität wird mit dem Programm *CHARMM* in einem *NVT*-Ensemble durchgeführt, d.h. Teilchenanzahl, Temperatur und Volumen bleiben konstant. Für jede Kavität werden Simulationen mit einer variierenden

Anzahl an Wassermolekülen innerhalb der Kavität durchgeführt und mit Hilfe der *Free Energy Perturbation* Methode (*FEP*-Methode) aus *CHARMM* die jeweilige Änderung der Freien Exzeß-Energie A bestimmt. Nach Glättung und Interpolation der Daten läßt sich durch Differentiation nach der verwendeten Wasseranzahl das Exzeß-chemische Potential m in Abhängigkeit der Wasseranzahl in der Kavität berechnen. Unter der Annahme, daß die Wassermoleküle der Kavität mit denen der *Bulk*-Phase im chemischen Gleichgewicht stehen, liegt eine optimale Wasserbesetzung dann vor, wenn das chemische Potential dem von reinem *Bulk*-Wasser (-6.4 kcal/mol) [119] entspricht (Abb. 3.2). Dabei können sich auch nicht-ganzzahlige Wasseranzahlen ergeben. Der Fehler der Methode liegt bei ungefähr $\Delta N = \pm 1$ Wassermolekül für jede Kavität.

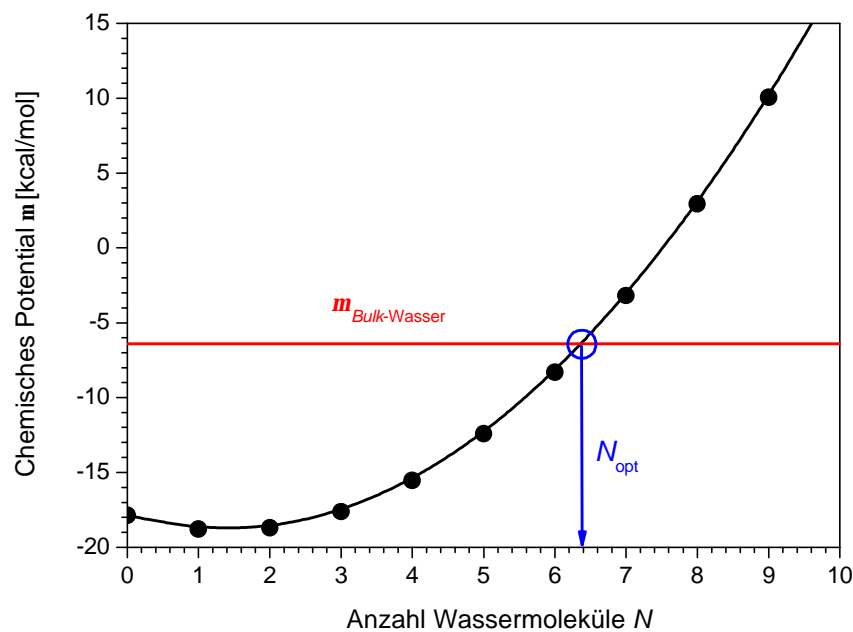


Abb. 3.2: Auftragung des chemischen Potentials m in Abhängigkeit der Wasseranzahl N innerhalb einer Kavität des Komplexes 1vfb aus dem Datensatz. Das chemische Potential wird durch MD-Simulationen und *FEP*-Berechnungen ermittelt. Unter der Annahme, daß die Wassermoleküle der Kavität mit denen der *Bulk*-Phase im chemischen Gleichgewicht stehen, ist die optimale Wasseranzahl innerhalb der Kavität erreicht, wenn m dem chemischen Potential von *Bulk*-Wasser entspricht.

Die durch *FEP*-Berechnungen ermittelte Anzahl an Wassermolekülen in den Kavitäten kann nun mit der Anzahl der darin befindlichen experimentellen Kristallwassermoleküle verglichen werden. Dazu werden jeder Kavität diejenigen Kristallwassermoleküle des Komplexes zugeordnet, die nicht weiter als 0.3 \AA von den Gitterpunkten, welche die Kavität bilden, entfernt sind.

Aus dem Vergleich der experimentellen und berechneten Wasseranzahl (Abb. 3.3) wird deutlich, daß in vielen Kavitäten keine experimentell ermittelten Kristallwassermoleküle vorliegen. Für größere Kavitäten werden weniger Kristallwassermoleküle beobachtet als berechnet werden. Dies ist auf die hohe Mobilität der Wassermoleküle in großen Kavitäten zurückzuführen, wodurch diese Moleküle in Röntgenstrukturanalysen nicht ermittelt werden können.

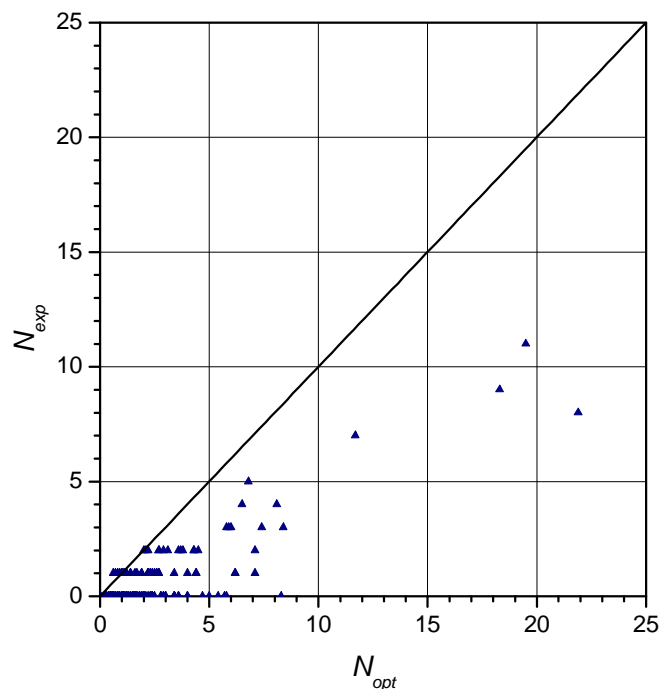


Abb. 3.3: Auftragung der experimentellen Kristallwasseranzahl (N_{exp}) einzelner Kavitäten gegen die optimale Wasseranzahl des Datensatzes aus *FEP*-Berechnungen (N_{opt}).

PDB-ID	Anzahl Cluster	H ₂ O Bereich	PDB-ID	Anzahl Cluster	H ₂ O Bereich	PDB-ID	Anzahl Cluster	H ₂ O Bereich
1a2y	6	0.5-6.8	1fns	5	1.1-4.3	1ncb	11	0.4-3.4
1ar1	9	0.5-4.3	1igc	9	0.0-4.0	1ncd	13	0.0-2.8
1bog	6	1.0-4.7	1jrh	6	0.7-5.8	1oak	6	0.0-3.6
1bql	11	0.9-8.4	1kip	5	1.0-21.9	1osp	8	0.0-11.7
1cic	7	0.5-7.4	1kiq	10	0.0-5.9	1qfu	11	0.0-5.4
1cu4	3	1.0-6.2	1kir	6	0.9-18.3	1vfb	5	0.9-19.5
1dqj	13	0.5-3.6	1mlc_1	9	0.2-2.0	2iff	11	0.4-8.3
1ejo	5	1.2-2.7	1mlc_2	8	0.0-2.1	3hfm	9	0.0-5.7
1eo8	9	0.3-2.9	1nca	10	0.0-2.9			

Tab. 3.2: Reduzierter Datensatz aus 26 Antikörper-Komplexen, für den die optimale Anzahl an Wassermolekülen in 211 Kavitäten (Clustern) durch MD-Simulationen und *FEP*-Berechnungen ermittelt wurde.

3.4.2 Optimierung des Radius der „Wasserkugel“

Die Optimierung des Radius der Wasserkugel erfolgt mit Hilfe des reduzierten Datensatzes (Tab. 3.2) an Proteinkomplexen, in dem für alle Kavitäten wie oben beschrieben durch MD-Simulationen die optimale Wasseranzahl berechnet wurde. Zur Optimierung des Radius wird das Programm *space* verwendet und das Verfahren zur Vorhersage der Wassermoleküle wie unter Abschnitt 3.3 beschrieben auf die Komplexe des reduzierten Datensatzes angewendet. Dabei wird in mehreren Durchgängen der „Wasserradius“ variiert, der den Abrollradius für die Generierung der Connolly-Oberflächen der beiden Komplexpartner sowie den Radius der Probekugel zum Auffüllen der Kavitäten beschreibt. Anschließend wird für jede Kavität die Anzahl der mit unterschiedlichen Radien berechneten Wassermoleküle mit der Anzahl der optimalen Wassermoleküle des Datensatzes verglichen. Durch die Variation des Wasserradius stimmen die neu berechneten Kavitäten geometrisch nicht exakt mit denen für die MD-Simulationen verwendeten Kavitäten überein. Daher werden letztere als Referenz genommen und jede neu berechnete Wasserposition derjenigen Referenzkavität zugeordnet, die nicht weiter als 0.3 Å entfernt liegt. Der optimale Wasserradius für das Verfahren ist dann erreicht, wenn die Abweichung der damit berechneten Wasseranzahl von der optimalen Wasseranzahl der MD-Simulationen für alle Kavitäten am geringsten ist. Dazu wird für jeden getesteten Radius die Wurzel der mittleren quadratischen Abweichung (*Root Mean Square Deviation* oder *RMS*-Wert) ermittelt:

$$RMS = \sqrt{\frac{1}{n} \cdot \sum_i (N_{opt,i} - N_{space,i})^2} \quad (3.1)$$

mit n : Anzahl der Kavitäten (Cluster),
 $N_{opt,i}$: optimale Wasseranzahl in der Kavität i aus dem Datensatz,
 $N_{space,i}$: mit *space* berechnete Wasseranzahl in der Kavität i .

Die *RMS*-Werte der unterschiedlichen Wasserradien zeigen, daß bei einem Wert von 1.3 Å die geringste Abweichung der mit *space* berechneten Wasseranzahl von der optimalen Wasseranzahl vorliegt (Tab. 3.3). Der entsprechende *RMS*-Wert von 0.93 liegt sogar noch innerhalb des Fehlerbereichs der durch *FEP*-Berechnungen bestimmten optimalen Wasseranzahl. Der Radius von 1.3 Å stimmt sehr gut mit Werten anderer Autoren überein,

die bei ähnlichen Verfahren zur Identifikation von Kavitäten im Innern von Proteinen einen Probekugelradius zwischen 1.2-1.3 Å als optimal ermittelten [59,61].

Wasserradius [Å]	1.2	1.3	1.4	1.5
RMS-Wert	1.05	0.93	1.31	1.71

Tab. 3.3: RMS-Werte der mit unterschiedlichen Wasserradien von *space* berechneten Wasseranzahl zur optimalen Wasseranzahl in den Kavitäten des Datensatzes.

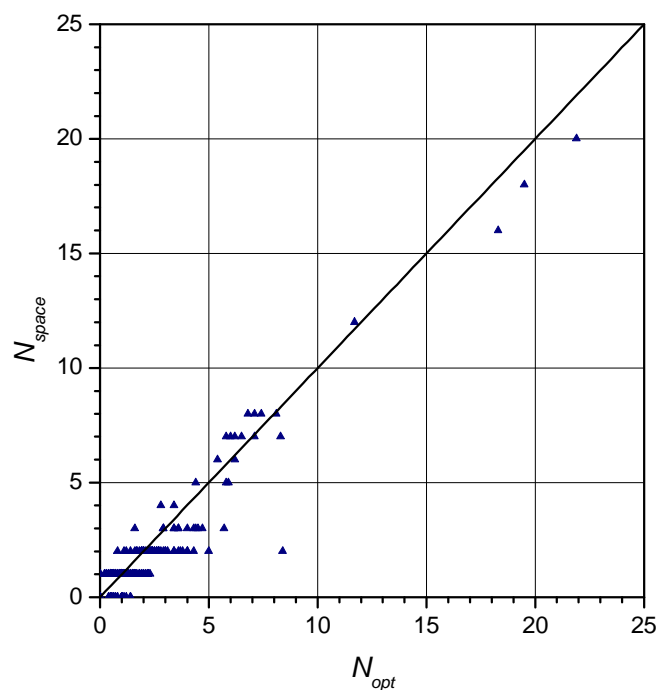


Abb. 3.4: Auftragung der durch *space* berechneten Wasseranzahl (N_{space}) einzelner Kavitäten bei einem Wasserradius von 1.3 Å gegen die optimale Wasseranzahl des Datensatzes aus *FEP*-Berechnungen (N_{opt}).

Die Auftragung der durch *space* berechneten Wasseranzahl mit einem Wasserradius von 1.3 Å gegen die optimale Wasseranzahl zeigt eine gute Übereinstimmung der Werte (Abb. 3.4). Größere Abweichungen sind auf die unterschiedlichen „inneren“ Radien bei der Erzeugung der Oberflächen der Komplexparker zurückzuführen. Dadurch können sich die Kavitäten, die für die *FEP*-Berechnungen erzeugt wurden, von denen mit variierenden Radien in ihrer Größe unterscheiden, was dann auch zu einer unterschiedlichen Wasseranzahl führt. Bei dem Antikörper des Komplexes 1bql beispielsweise führt ein schmaler Kanal von einer Kavität innerhalb der Binderegion tief ins Innere des Antikörpers

und verbindet zwei größere Hohlräume miteinander, die bei Verwendung eines Abrollradius von 1.2 \AA (wie bei der *FEP*-Berechnung verwendet) als eine Kavität außerhalb des Proteins angesehen werden (Abb. 3.5-A). Bei Verwendung eines größeren Abrollradius ist dieser Kanal allerdings verschlossen und führt nicht mehr zu dem dahinter liegenden Hohlraum, so daß dieser als Proteininneres definiert und bei der Berechnung verworfen wird (Abb. 3.5-B). Dies führt dazu, daß in diesem Fall nur 2 Wassermoleküle durch *space* vorhergesagt werden, obwohl bei der *FEP*-Berechnung 8 Wassermoleküle als optimale Besetzung vorliegen.

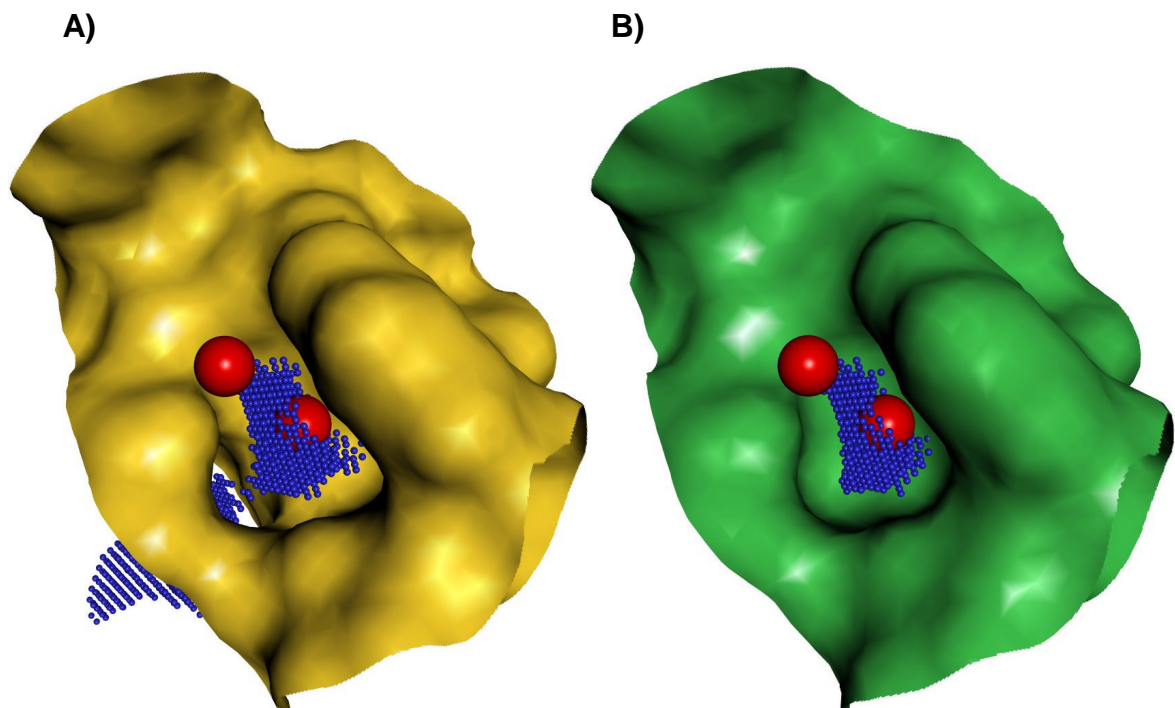


Abb. 3.5: Darstellung einer Kavität in der Binderegion des Antikörperkomplexes 1bql. Der Oberflächenbereich des Antikörpers, der die Kavität bildet, ist in gelb bzw. grün dargestellt. Die blauen Kugeln beschreiben die Geometrie der Kavität. Rote Kugeln entsprechen Sauerstoffatomen von Wassermolekülen und stellen eine mögliche Wasserplatzierung nach dem Auffüllvorgang durch das Programm *space* mit dem Wasserradius von 1.3 \AA dar. **A)** Wird die Oberfläche mit einem kleinen Abrollradius erzeugt (zur Verdeutlichung wurde hier 1.0 \AA verwendet), führt ein Kanal in das Protein, wo sich ein weiterer Hohlraum befindet, so daß eine sehr viel größere Kavität resultiert. **B)** Bei Verwendung eines größeren Abrollradius wird der Kanal geschlossen; der zweite Hohlraum befindet sich innerhalb des Proteins und wird bei der Wasserberechnung nicht berücksichtigt.

3.4.3 Einfluß des „äußeren“ Radius

Durch den Radius der Abrollkugel zur Generierung der Connolly-Oberfläche des Gesamtkomplexes, „äußerer“ Radius genannt, wird festgelegt, ob Kanäle, die in die Binderegion des Komplexes führen und Kavitäten mit dem *Bulk*-Wasser verbinden, durch die Komplexoberfläche bedeckt werden oder offen sind. Wird der Kanal durch die Komplexoberfläche nicht geschlossen, liegen die Punkte, die den Kanal und die mit ihm verbundene Kavität definieren, außerhalb der Komplexoberfläche und werden daher als *Bulk*-Region klassifiziert. Ist der Kanal hingegen geschlossen, befinden sich diese Punkte innerhalb der Komplexoberfläche und werden als Kavität erkannt. Durch die Wahl des äußeren Radius bleibt es also dem Benutzer überlassen, ab welchem Kanaldurchmesser die betroffenen Regionen als Kavität bzw. *Bulk*-Region definiert werden. In dieser Arbeit wird als äußerer Radius ein Wert von 1.8 Å gewählt. Somit darf die Öffnung eines potentiellen Kanals in die Binderegion nur geringfügig größer als ein Wassermolekül sein, ansonsten wird die damit verbundene Region nicht mehr als Kavität erkannt. Um den Einfluß des äußeren Radius auf die Vorhersage von Kavitäten und Wassermolekülen in der Binderegion zu testen, wird dieser Radius für die Berechnung der Kavitäten mit *space* variiert. Als Datensatz wird der in Tab. 3.1 aufgeführte Gesamtdatensatz von 111 Proteinkomplexen verwendet. Der innere Radius beträgt bei den Berechnungen 1.3 Å.

Die Ergebnisse der Berechnungen sind in den Abbildungen 3.6-3.9 dargestellt. Mit wachsendem äußeren Radius nimmt die Anzahl der identifizierten Kavitäten zunächst zu (Abb. 3.6). Dies liegt daran, daß immer breitere Kanäle durch die Komplexoberfläche abgedeckt werden und diese Regionen als Kavitäten klassifiziert werden. Das Maximum liegt bei ca. 3 Å. Wird der Radius noch weiter vergrößert, sinkt die Anzahl der Kavitäten wieder. Dies bedeutet, daß sich bei weiterer Erhöhung des Abrollradius räumlich getrennte Kavitäten am Rand der Binderegion vergrößern und sich schließlich so nahe kommen, daß sie eine Kavität bilden und die Gesamtzahl daher abnimmt. Die Kavitäten können sich nur am Rand vergrößern und „wachsen“ immer weiter von der Binderegion weg, da durch den größeren Abrollradius Spalten an der Außenseite der Proteine abgedeckt werden, die mit der Kavität in der Binderegion in Verbindung stehen und daher nicht mehr verworfen werden (Abb. 3.7).

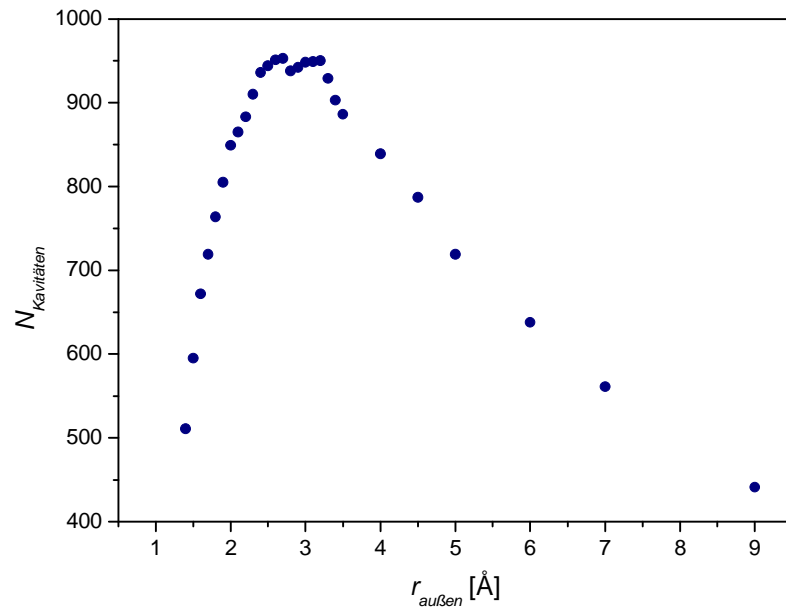


Abb. 3.6: Anzahl der Kavitäten $N_{\text{Kavitäten}}$ in Abhängigkeit vom äußeren Radius $r_{\text{außen}}$.

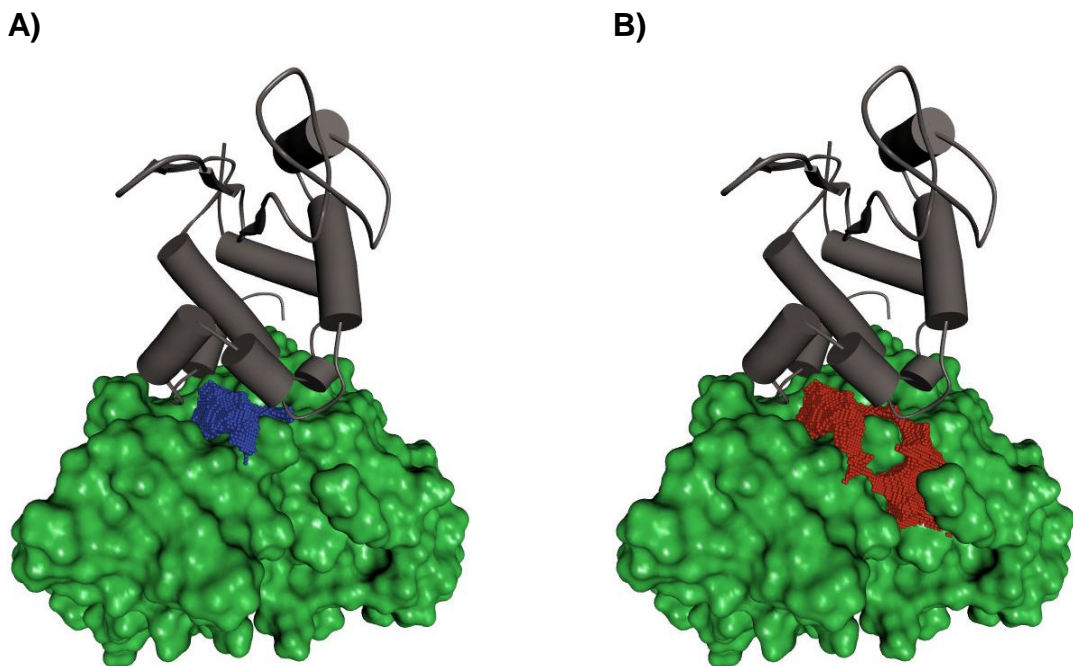


Abb. 3.7: Darstellung des Antikörperkomplexes 1vfb mit einer ausgewählten Kavität in der Binderegion. Die Connolly-Oberfläche des F_v -Fragmentes ist in grün dargestellt, der Komplexpartner ist durch ein graues *Ribbon* wiedergegeben. Zur Verdeutlichung ist nur eine Kavität der Binderegion dargestellt. Sie wird durch Gitterpunkte beschrieben, die als kleine Kugeln (blau bzw. rot) dargestellt sind. **A)** Die blaue Kavität wurde mit einem äußeren Radius von 1.8 Å berechnet. Wird dieser Radius erhöht, werden Spalten des Proteins in der Nähe der Kavität durch den größeren Abrollradius bedeckt. Sie befinden sich dann innerhalb der Komplexoberfläche, was dazu führen kann, daß die Kavität bis in diese neuen Bereiche „hineinwächst“. **B)** Bei Verwendung eines äußeren Radius von 3.5 Å wird deutlich, daß sich die neue Kavität (rot) im Vergleich zur ursprünglichen in die benachbarten Spalten ausbreitet.

Obwohl die Anzahl der Kavitäten mit wachsendem äußeren Radius abnimmt, werden trotzdem auch neue Kavitäten identifiziert. Eine neue Kavität liegt dann vor, wenn alle ihre Gitterpunkte weiter als 2.6 \AA von den Kavitäten des nächstkleineren äußeren Radius entfernt sind. Abb. 3.8 zeigt, daß die Anzahl neuer Kavitäten mit wachsendem Radius stetig abnimmt, so daß allmählich alle Kanäle zugedeckt werden. Bei einem Radius von 4 \AA und 6 \AA steigt die Anzahl an neuen Kavitäten plötzlich wieder. Möglicherweise beschreiben diese Radien charakteristische Größen von Kanälen, die alle bei einer solchen Probekugelgröße verschlossen werden. Die Anzahl der Wassermoleküle nimmt im beobachteten Wertebereich des Radius ($1.4\text{-}9 \text{ \AA}$) stetig zu (Abb. 3.9). Dies stützt die These, daß sich die Kavitäten bei abnehmender Anzahl vergrößern und zusammenschließen müssen, um die wachsende Wasseranzahl aufnehmen zu können.

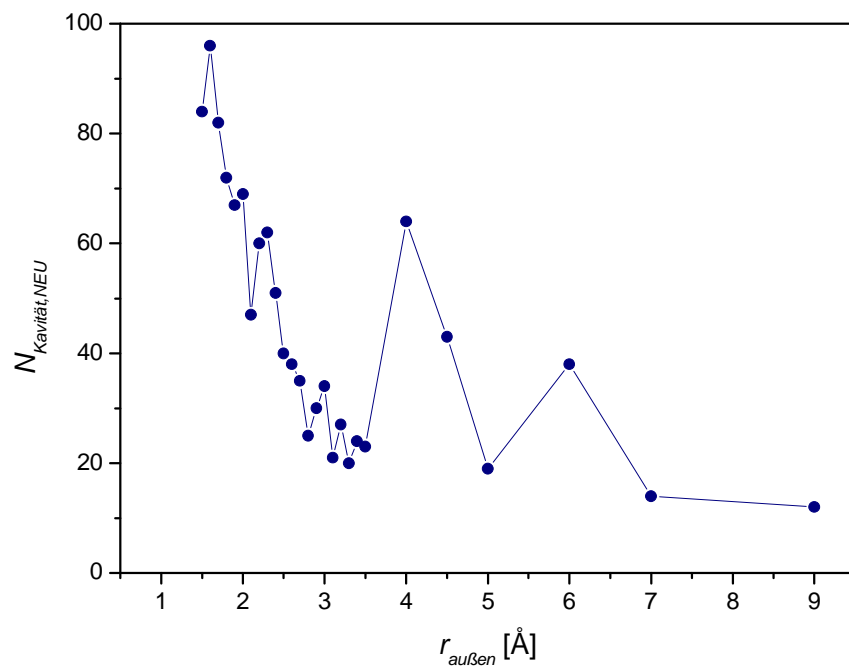


Abb. 3.8: Anzahl der neuen Kavitäten $N_{\text{Kavitäten,NEU}}$ in Abhängigkeit vom äußeren Radius $r_{\text{außen}}$.

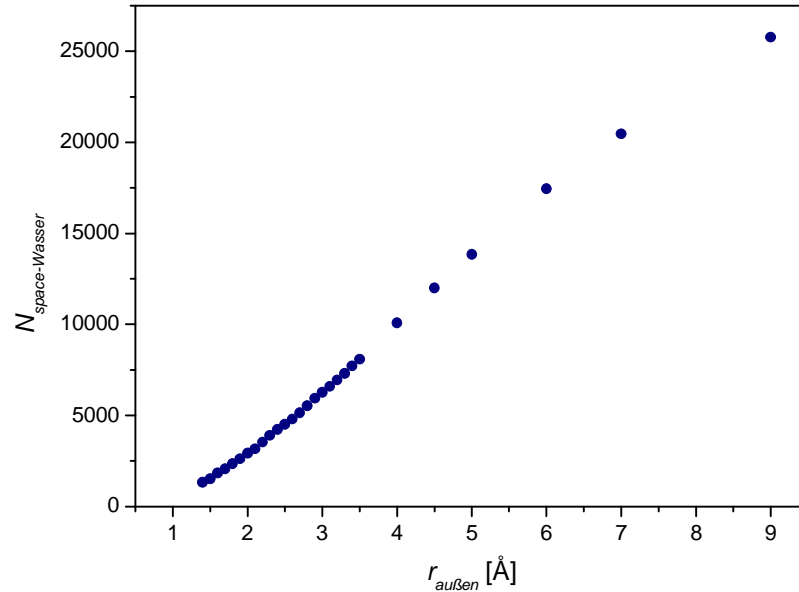


Abb. 3.9: Abhängigkeit der mit *space* berechneten Wasseranzahl $N_{\text{space-Wasser}}$ vom äußeren Radius $r_{\text{außen}}$.

3.5 Optimierung des Auffüllvorganges mit der Erweiterung *sphopt*

Nach erfolgter Identifikation und Einteilung in räumlich getrennte Cluster werden die Kavitäten in dem Programm *space* nach einem deterministischen, geometrischen Algorithmus mit Kugeln aufgefüllt. Dabei wird nicht zwangsläufig die optimale Kugelpackung erreicht. Das Verfahren wird daher durch das Programm *sphopt* (O. Korb & J. Gimmler, persönliche Mitteilung) erweitert, mit dem eine Optimierung der Packungsdichte erreicht werden soll. Als Eingabe werden nur die von *space* erzeugten Gitterpunkte benötigt, die einen Cluster darstellen. Das Programm beruht auf einem iterativen, heuristischen Verfahren (*Iterated-Greedy-Verfahren*) [120]. Dabei wird zunächst eine mögliche Startlösung aufgebaut, die nicht dem Optimum entsprechen muß. In dem vorliegenden Beispiel werden die von *space* berechneten Wasserpositionen als Startkonfiguration verwendet. Aus dieser Startkonfiguration wird zufällig eine Wasserkugel ausgewählt und alle benachbarten Kugeln innerhalb eines gewissen *Cutoff*-Radius (5 Å) herausgelöscht. Der so entstandene Freiraum wird anschließend durch zufällige Auswahl von Gitterpositionen wieder aufgefüllt. Erhöht sich dadurch die Anzahl der Wasserkugeln, oder werden bei gleicher Anzahl weniger Gitterpunkte des Clusters bedeckt, wird die neue Lösung übernommen. Ausgehend von der entstandenen Konfiguration wird eine neue Lösung aufgebaut. Durch Iteration (50000 Schritte) des Prozesses wird die Lösung sukzessive verbessert.

Zum Vergleich der Ergebnisse von *space* und seiner Erweiterung *sphopt* werden die in Abschnitt 3.4.2 mit *space* bestimmten Kavitäten unterschiedlicher Wasserradien verwendet und mit *sphopt* neu aufgefüllt. Anschließend wird der *RMS*-Wert (Gl. 3.1) der so erhaltenen Wasseranzahl in Bezug auf die optimale Wasseranzahl der *FEP*-Berechnung ermittelt (Tab. 3.4).

Wasserradius [Å]	1.2	1.3	1.4	1.5
<i>RMS</i> -Wert	1.50	0.96	1.19	1.65

Tab. 3.4: *RMS*-Werte der mit unterschiedlichen Wasserradien von *space* und *sphopt* berechneten Wasseranzahl zur optimalen Wasseranzahl in den Kavitäten des Datensatzes.

Der beste *RMS*-Wert liegt, wie auch schon bei der Verwendung von *space* ohne die *sphopt*-Erweiterung, bei einem Wasserradius von 1.3 Å. Die Erweiterung bringt allerdings für die Vorhersagegenauigkeit keine Verbesserung, da der *RMS*-Wert von 0.96 fast mit dem des Verfahrens ohne Erweiterung (0.93) identisch ist.

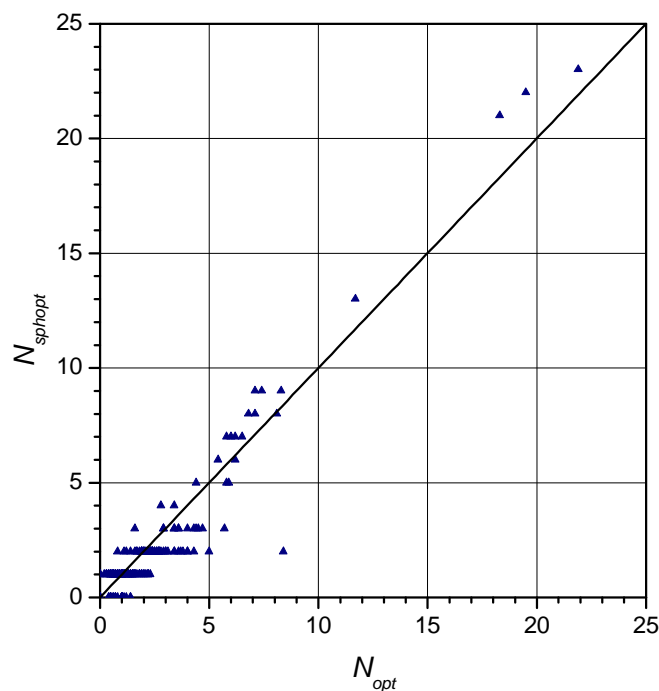


Abb. 3.10: Auftragung der durch *space* und *sphopt* berechneten Wasseranzahl (N_{sphopt}) einzelner Kavitäten bei einem Wasserradius von 1.3 Å gegen die optimale Wasseranzahl aus *FEP*-Berechnungen (N_{opt}).

Die Auftragung der mit *sphopt* berechneten Wasseranzahl gegen die optimale Wasseranzahl aus dem Datensatz ähnelt der des reinen *space*-Verfahrens und zeigt ebenfalls eine gute Anpassung an die optimalen Werte (Abb. 3.10). Es fällt allerdings auf, daß bei großen Kavitäten die Wasseranzahl durch die Optimierung der Packungsdichte zu groß wird, während bei dem reinen *space*-Verfahren die Wasseranzahl im Vergleich zur optimalen Besetzung zu klein ist. Da durch die Erweiterung zusätzliche Rechenzeit in Anspruch genommen wird, das Verfahren jedoch keine Verbesserung in der Vorhersagegenauigkeit der Wasseranzahl birgt, kann je nach Anwendung bzw. Größe der Kavitäten auf die *sphopt*-Erweiterung verzichtet werden.

3.6 Vergleich des Verfahrens mit dem Programm *DOWSER*

Das von Hermans *et al.* [57] entwickelte Programm *DOWSER* ermöglicht ebenfalls die Berechnung von Wassermolekülen in Kavitäten und Spalten von Proteinen. Im Unterschied zu dem Programm *space* wird hier eine Energieminimierung durchgeführt, um die optimalen Wasserpositionen zu berechnen. Allerdings gibt es keine Möglichkeit, nur die Wasserpositionen innerhalb der Binderegion zu bestimmen. Die Ausgabe enthält sehr viele Wassermoleküle, die sich innerhalb der einzelnen Proteinpartner oder in Spalten an der Außenseite der Proteine befinden, die jedoch nicht zur Binderegion gehören und daher später aussortiert werden müssen. Mit dem Programm-Modul *dowserx* werden in einem ersten Durchgang unter Verwendung der Standardparameter die Wasserpositionen bestimmt. Als Eingabedatei werden die Koordinaten des gesamten Proteinkomplexes ohne die Kristallwassermoleküle verwendet. Hierdurch werden meist nur Wassermoleküle identifiziert, die in direktem Kontakt mit dem Protein stehen, so daß größere Kavitäten nicht vollständig ausgefüllt werden. In einem zweiten Schritt wird das Verfahren daher mit dem Modul *dowser-repeat* wiederholt, wobei die identifizierten Wassermoleküle des ersten Schrittes als Teil des Proteinkomplexes angesehen und bei der erneuten Berechnung berücksichtigt werden. Aus den mit dem beschriebenen Verfahren berechneten Wassermolekülen werden anschließend diejenigen Wassermoleküle ausgewählt, die nicht weiter als 0.3 Å von den Gitterpunkten der Kavitäten aus dem optimalen Datensatz entfernt sind.

Grundlage für den Vergleich der Ergebnisse der beiden Programme *space* und *DOWSER* ist wieder der in Abschnitt 3.4.1 vorgestellte reduzierte Datensatz an Proteinkomplexen. Der *RMS*-Wert der mit *DOWSER* bestimmten Wasseranzahl in Bezug auf die optimalen Werte der *FEP*-Berechnungen liegt mit 1.96 deutlich über dem *RMS*-Wert des *space*-Verfahrens von 0.93. Die gute Vorhersagegenauigkeit der Wasseranzahl durch *space* wird mit dem *DOWSER*-Programm keineswegs erreicht. Die starke Abweichung der mit *DOWSER* berechneten Wasseranzahl von den optimalen Werten wird auch an der Auftragung der Werte in Abbildung 3.11 deutlich.

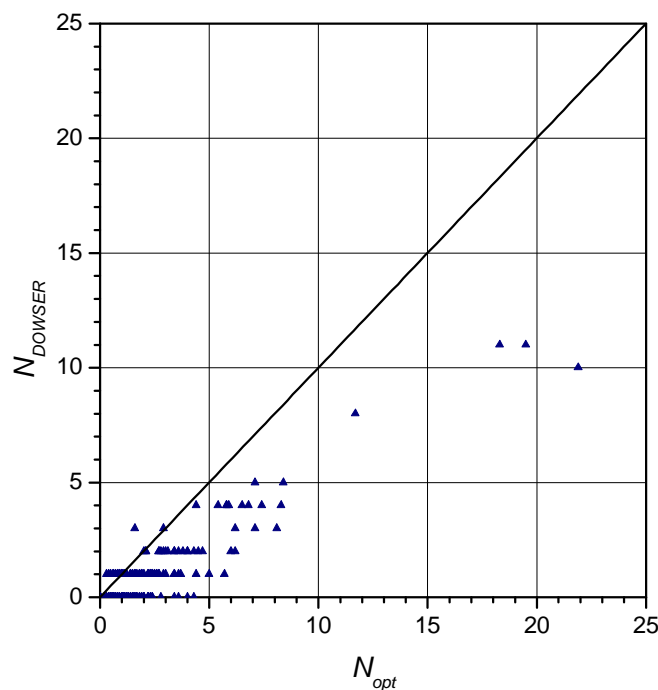


Abb. 3.11: Auftragung der durch *DOWSER* berechneten Wasseranzahl (N_{DOWSER}) einzelner Kavitäten gegen die optimale Wasseranzahl aus *FEP*-Berechnungen (N_{opt}).

Aus den Ergebnissen kann gefolgert werden, daß für die Analyse von Wassermolekülen in der Binderegion von Proteinkomplexen das in dieser Arbeit entwickelte Programm *space* besser geeignet ist als *DOWSER*. Der große Vorteil von *space* liegt in der automatisierten Identifikation von Kavitäten in der Binderegion und nicht in der Peripherie oder im Innern der Proteine. Darüber hinaus wird die Wasseranzahl in den Kavitäten genauer vorhergesagt. Daher wird das Programm *space* für die statistischen Untersuchungen der Proteinkomplexe in Kapitel 4 herangezogen.

3.7 Validierung des Verfahrens mit Kristallwassermolekülen

In den vorhergehenden Abschnitten wurde das *space*-Verfahren dahingehend untersucht, wie genau die Wasseranzahl innerhalb der vorgeschlagenen Kavitäten vorhergesagt werden kann. Im folgenden soll nun geprüft werden, wie viele der experimentell beobachteten Kristallwassermoleküle der Binderegion sich in den berechneten Kavitäten befinden, und ob es Kristallwassermoleküle gibt, deren Position das Verfahren nicht vorhersagen kann. Dazu mußten zunächst diejenigen Kristallwassermoleküle identifiziert werden, die innerhalb der Binderegion zwischen den beiden Proteinkomplexen liegen. Da die räumliche Begrenzung der Binderegion durch das *space*-Verfahren definiert ist, mußte für die Identifikation der experimentellen Kristallwassermoleküle ein entsprechendes Verfahren entwickelt werden, das nur Kristallwassermoleküle in dieser definierten Region findet.

Das Verfahren zur Identifikation der Kristallwassermoleküle in der Binderegion ähnelt der Methode aus Abschnitt 3.3, die zur Berechnung der Kavitäten angewendet wird. Zunächst wird für jeden der beiden Komplexpartner die jeweilige Connolly-Oberfläche mit einem Abrollradius von 1.3 Å sowie die Connolly-Oberfläche des Gesamtkomplexes mit einem Abrollradius von 1.8 Å bestimmt. Alle Kristallwassermoleküle, von denen ein Atom innerhalb der Komplexoberfläche und außerhalb der Oberflächen der einzelnen Komplexpartner sitzt, werden ausgewählt. Die restlichen Kristallwassermoleküle werden verworfen. Befinden sich Kristallwassermoleküle der Auswahl weiter als 6 Å von einem der beiden Komplexpartner entfernt, oder befindet sich das Sauerstoffatom eines Wassermoleküls außerhalb der Komplexoberfläche, werden diese Wassermoleküle ebenfalls entfernt. Die verbliebene Auswahl stellt die Kristallwassermoleküle der Binderegion dar.

Die so bestimmten Kristallwassermoleküle können nun zur Überprüfung des *space*-Programmes verwendet werden. Die Überprüfung erfolgt mit dem Gesamtdatensatz aus 111 Proteinkomplexen aus Tab. 3.1. Zunächst werden für den Datensatz die Kavitäten mit dem *space*-Verfahren (Abschnitt 3.3) berechnet und anschließend die Kristallwassermoleküle der Binderegion identifiziert. Ist ein Kristallwassermolekül nicht weiter als 0.8 Å von einem der Gitterpunkte entfernt, die eine Kavität beschreiben, wird es dieser Kavität zugeordnet. Kristallwassermoleküle, die keiner Kavität zugeordnet werden können, beschreiben Positionen, bei denen die Kavitäts-Erkennung des *space*-Verfahrens versagt.

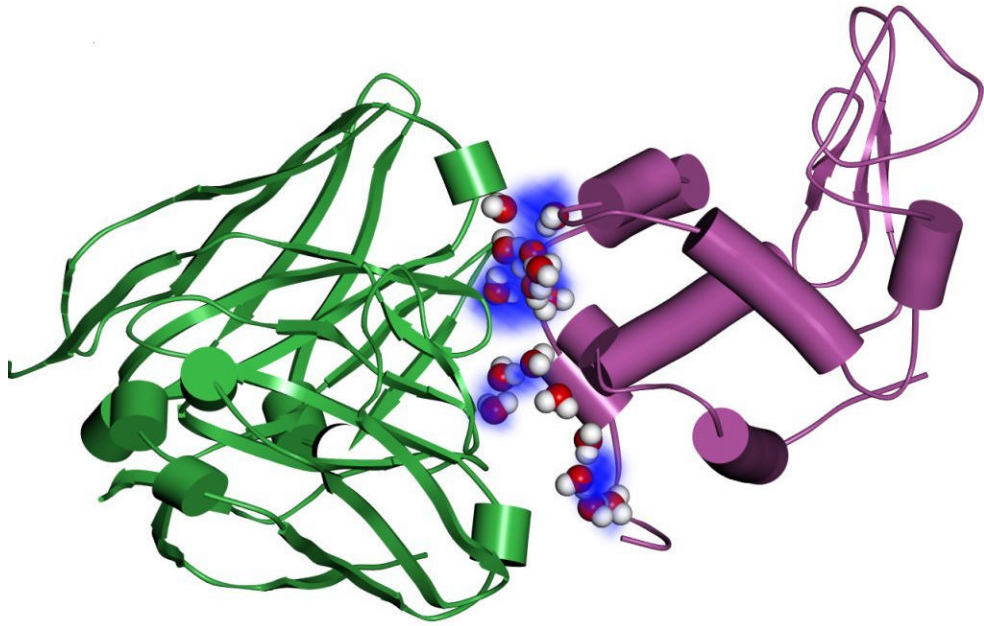


Abb. 3.12: Komplex 1vfb aus dem Antikörperfragment F_v D1.3 (grün) und seinem Komplexpartner HE-Lysozym (magenta). Die blauen Wolken entsprechen den mit *space* berechneten Kavitäten und beschreiben mögliche Wasserpositionen in der Binderegion des Komplexes. Die experimentell bestimmten Kristallwassermoleküle der Binderegion liegen genau in den vorhergesagten Kavitäten.

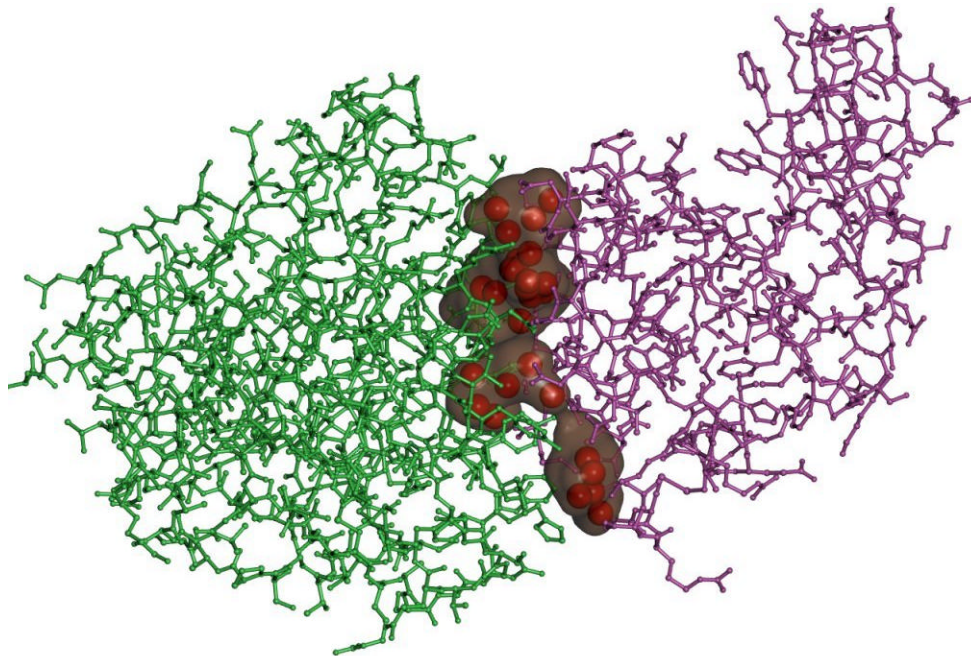


Abb. 3.13: Komplex und Orientierung entsprechen Abb. 3.12. Die mit *space* berechneten Kavitäten sind durch braune, transparente Oberflächen dargestellt. Die roten Kugeln innerhalb der Kavitäten sind die durch *space* platzierten „Wasserkugeln“, deren Positionen als Startkoordinaten für Sauerstoffatome beim Aufbau und der Energieminimierung von Wassermolekülen in der Binderegion dienen können, falls keine experimentellen Kristallwasserdaten vorliegen.

Von den insgesamt 799 identifizierten Kristallwassermolekülen in der Binderegion des Datensatzes liegen 760 in den mit *space* berechneten Kavitäten. Die Vorhersage potentieller Wasserpositionen mit *space* ist also als sehr gut zu bezeichnen, da 95% der experimentellen Kristallwassermoleküle auch tatsächlich innerhalb der berechneten Regionen zu finden sind. Abb. 3.12 und Abb. 3.13 zeigen an einem Beispielkomplex des Datensatzes die Positionen der mit *space* berechneten Kavitäten sowie die identifizierten Kristallwassermoleküle. Daß 5% der Kristallwassermoleküle Regionen besetzen, die nicht vorhergesagt werden, ist vor allem auf die Gitterauflösung zurückzuführen. Handelt es sich beispielsweise um ein einzelnes Wassermolekül, daß extrem eng von den benachbarten Proteinatomen umgeben ist, gibt es nur wenige Positionen, an denen eine Probekugel platziert werden kann, um in dieser Kavität Platz zu finden. Liegen diese Position nicht genau auf einem der Gitterpunkte, kann die Kavität nicht einfach identifiziert werden. Durch Verkleinerung des Gitterabstandes kann die Vorhersage demnach verbessert werden, der Rechenaufwand nimmt dadurch allerdings erheblich zu.

3.8 Diskussion der Ergebnisse des *space*-Verfahrens

Das vorgestellte Verfahren zur Berechnung von Kavitäten und potentiellen Wasserpositionen in der Binderegion von Proteinkomplexen zeigt nach der Optimierung seiner Parameter eine sehr gute Übereinstimmung mit den Daten eines aus MD-Simulationen und *FEP*-Berechnungen erzeugten Datensatzes an Proteinkomplexen. Bei einem Probekugelradius („Wasserradius“) von 1.3 Å zur Identifikation und Besetzung der Kavitäten mit Wasserkugeln beträgt die Abweichung (*RMS*-Wert) zur optimalen Wasseranzahl in den Kavitäten des Datensatzes nur 0.93. Die Größe dieser Probekugel ist entscheidend dafür, ob Kavitäten gefunden werden und wie viele Wasserkugeln darin platziert werden können. Hubbard *et al.* [59] entwickelten einen Algorithmus zur Identifikation von Kavitäten im Innern globulärer Proteine, der ebenfalls auf einem gitterbasierten Verfahren mit Probekugeln beruht. Der optimale Proberadius zur Identifikation liegt bei diesem Verfahren bei 1.25 Å. Auch von Poornima & Dean [61] werden optimale Werte zwischen 1.2-1.4 Å für den Proberadius diskutiert. Diese Werte stimmen sehr gut mit dem in dieser Arbeit bestimmten optimalen Proberadius von 1.3 Å überein. Das *space*-Verfahren weist außerdem eine gute Reproduktion der Positionen experimenteller Kristallwassermoleküle der Binderegion auf, da 95% dieser Kristallwasser-

moleküle innerhalb der vorhergesagten Kavitäten liegen. Im Vergleich zu anderen Arbeiten [51-56], mit denen ebenfalls Kavitäten in Proteinen berechnet werden können, hat das *space*-Verfahren für die Analyse von Protein-Protein-Wechselwirkungen zwei entscheidende Vorteile. Zum einen beschränkt sich der Großteil der Verfahren auf Kavitäten innerhalb des Proteins und beinhaltet keine Selektion von Kavitäten zwischen den Komplexpartnern, also innerhalb der Binderegion des Komplexes. Eine Ausnahme bildet hier das von Laskowski *et al.* [55] entwickelte Programm *SURFNET*, mit dem sogenannte *Gaps* („Lücken“) zwischen Komplexpartnern berechnet werden können. Das *space*-Verfahren liefert automatisch nur die in der Binderegion liegenden Kavitäten, obwohl die Bestimmung von Kavitäten innerhalb der Proteine durch leichte Modifikation des Verfahrens prinzipiell auch möglich wäre. Ein weiterer Vorteil des hier entwickelten Verfahrens liegt in der Vorhersage der Anzahl und Position von Wassermolekülen innerhalb der Kavitäten. Außer bei dem Programm *DOWSER* [57] können die Kavitäten nicht mit Wassermolekülen gefüllt werden. Im Vergleich mit dem *space*-Verfahren schneidet das *DOWSER*-Programm in der Vorhersage der Wasseranzahl der Kavitäten mit einem *RMS*-Wert von 1.96 bezogen auf die MD-Simulationsergebnisse allerdings deutlich schlechter ab. Hinzu kommt, daß bei *DOWSER* nicht zwischen Kavitäten der Binderegion und dem Proteininnern unterschieden wird. Für die in dieser Arbeit durchgeführte Analyse von Wassermolekülen in der Komplexbinderegion von Protein-Protein-Komplexen stellt das *space*-Verfahren daher die beste Methode dar.

4 Statistische Analyse der Binderegion

Das in Kapitel 3 entwickelte Programm *space* stellt ein hilfreiches Werkzeug zur Analyse der Binderegion von Protein-Protein-Komplexen dar. Mit diesem Verfahren ist es nun möglich, eine Vielzahl von Komplexen automatisiert nach dem Vorhandensein von Kavitäten und möglichen Wasserpositionen in der Binderegion zu untersuchen, ohne auf Daten experimentell ermittelter Kristallwassermoleküle angewiesen zu sein. Als Datenbasis für alle folgenden Untersuchungen an Protein-Protein-Komplexen dient der in Tab. 3.1 zusammengestellte Datensatz. Anhand dieses Datensatzes sollen Anzahl, Verteilung und Packungsdichte der mit *space* berechneten Wassermoleküle in der Komplexbinderegion beschrieben werden. Darüber hinaus wird untersucht, ob sich die Oberflächenbereiche, die Protein-Protein-Kontakte ausbilden, von Bereichen, die eine mit Wasser gefüllte Kavität ausbilden, hinsichtlich ihrer geometrischen und physikochemischen Eigenschaften unterscheiden. Im folgenden werden die Methoden und Ergebnisse dieser Analysen näher erläutert.

4.1 Berechnung physikochemischer und geometrischer Daten

Zur Analyse der Komplexbinderegionen werden wie in Abschnitt 3.3 beschrieben mit dem *space*-Verfahren zunächst die molekularen Oberflächen, die Kavitäten sowie die Wasseranzahl für jeden Komplex aus den aufgearbeiteten Kristallstrukturen des gesamten Datensatzes berechnet. Anschließend werden wie unten beschrieben die Oberflächenanteile der Binderegion bestimmt und in Protein-Protein- und Protein-Kavität-Kontaktflächen eingeteilt. Die Analyse der physikochemischen und strukturellen Eigenschaften in der Binderegion erfolgt dann durch Berechnung dieser Eigenschaften auf den molekularen Oberflächenanteilen der Komplexpartner, welche die Binderegion darstellen.

4.1.1 Bestimmung und Unterteilung der Oberflächenanteile der Binderegion

Die Berechnung der gesamten Binderegion und deren Unterteilung in zwei unterschiedliche Kontaktbereiche basiert auf Abstandsmessungen zwischen den Oberflächen der einzelnen Komplexpartner bzw. zwischen den Oberflächen und den eine Kavität bildenden Gitterpunkten. Die drei Teilbereiche sind folgendermaßen definiert:

- *Gesamte Binderegion:*
Oberflächenanteile, die nicht weiter als 1.8 \AA von der Oberfläche des Komplexpartners *oder* von einem Gitterpunkt einer Kavität entfernt sind.
- *Protein-Protein-Kontaktregion:*
Oberflächenanteile, die nicht weiter als 1.8 \AA von der Oberfläche des Komplexpartners *und* weiter als 1.8 \AA von allen Gitterpunkten der Kavitäten entfernt sind.
- *Protein-Kavität-Kontaktregion:*
Oberflächenanteile, die nicht weiter als 1.8 \AA von den Gitterpunkten einer Kavität entfernt sind.

Dabei stellen die Protein-Protein- und die Protein-Kavität-Kontaktregionen Teilmengen der gesamten Binderegion dar. Die Abstände der Oberflächen werden mit dem Programm *interfacing* [35], die Abstände zu den Gitterpunkten der Kavitäten mit dem Programm *pdbdist* (M. Keil, persönliche Mitteilung) berechnet. Auf diesen Berechnungen basierend weist das Programm *selectqual* jedem Oberflächenpunkt eine Eigenschaft zu, die codiert zu welchen Regionen er gehört. Danach wird die Binderegion mit dem Programm *cutsurf* (M. Keil, persönliche Mitteilung) ausgeschnitten, so daß für diese Teiloberfläche die physikochemischen Eigenschaften - wie unter Abschnitt 4.1.2 beschrieben - bestimmt werden können. Anschließend wird die Binderegion erneut mit *cutsurf* in die beiden unterschiedlichen Kontaktbereiche unterteilt. In Abbildung 4.1 ist die Unterteilung der Binderegion in unterschiedliche Kontaktbereiche am Beispiel des Antikörperfragmentes F_v D1.3 aus dem Komplex 1vfb gezeigt.

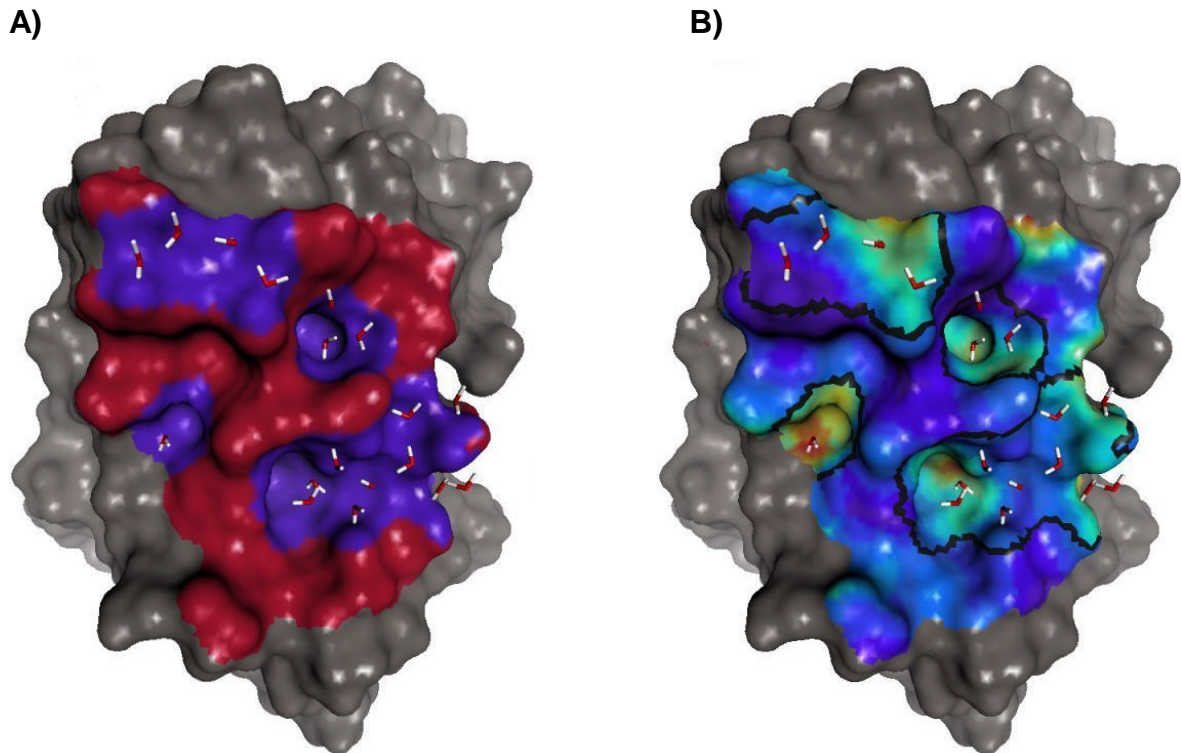


Abb. 4.1: Blick auf die Bindungsregion des Antikörperfragmentes F_v D1.3 aus dem Komplex 1vfb. Kristallwassermoleküle sind als *Sticks* dargestellt. **A)** Regionen mit direktem Kontakt zum Komplexpartner sind in rot dargestellt, Regionen, die Kavitäten bilden, sind blau. Graue Oberflächenanteile befinden sich außerhalb der Binderegion. **B)** Farbcodierte Darstellung der Dichte von Wasserstoffdonatoren und -akzeptoren (Zunahme von blau nach rot) in der Binderegion. Die unterschiedlichen Kontaktregionen sind zur Verdeutlichung durch schwarze Linien voneinander getrennt.

4.1.2 Berechnung physikochemischer/geometrischer Daten auf der Proteinoberfläche

Zur Analyse der Binderegion werden die folgenden geometrischen und physikochemischen Eigenschaften der Proteine berechnet und auf die molekulare Teiloberfläche projiziert, die die Binderegion darstellt: Elektrostatisches Potential (ESP), Molekulare Lipophilie (MLP), Verteilung von Wasserstoffakzeptoren/-donatoren (HB2), Salzbrückenverteilung (SB2), Tiefeninformation (CAV). Die Wasserstoffbrückenbindungsdichte (HD), der *Shapeindex* (SHI) sowie die *Curvedness* (CUR) werden im Unterschied zu den anderen Eigenschaften zunächst für die Gesamtoberfläche der einzelnen Proteine berechnet, erst dann wird die Binderegion mit *cutsurf* ausgeschnitten. Dies ist notwendig, da für die Berechnung dieser drei Eigenschaften auf einem Oberflächenpunkt alle benachbarten Punkte der Gesamtoberfläche berücksichtigt werden müssen. Im folgenden werden die Programme und Parameter zusammengestellt, die für die Berechnung der Eigenschaften auf den

Oberflächen verwendet werden. Sind bei der folgenden Zusammenstellung die Programmparameter nicht explizit aufgeführt, werden die vom Programm voreingestellten Standardwerte verwendet.

- ESP: Die Partialladungen der Atome basieren auf dem *CHARMM*-Kraftfeld und werden mit dem Programm *getesp* [77] aus der nach der Energieminimierung des Proteins von *CHARMM* erzeugten PSF-Datei (*Protein Structure File*) entnommen. Die Berechnung des elektrostatischen Potentials auf der Oberfläche erfolgt mit der in Abschnitt 2.5.1 beschriebenen Abstandsfunktion (Gl. 2.2) durch das Programm *qualen2* (*Cutoff* 11 Å, $\epsilon = \epsilon_0$) [35].
- MLP: Die Zuordnung der Lipophilie-Werte für jedes Atom nach Ghose und Crippen erfolgt mit dem Programm *newcripp* (M. Keil, persönliche Mitteilung). Die Berechnung der lokalen Lipophilie auf der Oberfläche entspricht der in Abschnitt 2.5.2 aufgeführten Abstandsfunktion (Gl. 2.6) und wird mit dem Programm *qualen2* (Parameter $d = 2 \cdot a = 2.0$, Parameter $c = b = 2.5$) durchgeführt.
- HB2: Die Einteilung der Atome in Wasserstoffdonatoren und -akzeptoren erfolgt mit dem Programm *newhbond* (M. Keil, persönliche Mitteilung) nach der in Abschnitt 2.5.3.1 beschriebenen Methode. Die Projektion auf die Moleküloberfläche wird mit dem Programm *hb* (M. Keil, persönliche Mitteilung) durchgeführt.
- SB2: Die Zuordnung der Ladung der Atome erfolgt mit dem Programm *gensb*. Dabei werden alle potentiell geladenen Aminosäureseitengruppen und Termini wie in Abschnitt 2.5.4 beschrieben als geladen behandelt. Die Projektion auf die Moleküloberfläche wird mit dem Programm *sb* durchgeführt.
- CAV: Die Tiefeninformation über Taschen und Spalten wird mit dem Programm *deepqual* [35] nach dem in Abschnitt 2.5.5.2 beschriebenen Verfahren berechnet. Der Radius der abdeckenden Probekugel beträgt dabei 6.0 Å.
- HD: Die Berechnung der kombinierten Dichte von Wasserstoffdonatoren und -akzeptoren erfolgt wie in Abschnitt 2.5.3.2 beschrieben mit dem Programm *hbdens* [35]. Damit der Einfluß der Umgebung auf die Dichte eng begrenzt bleibt, wird als *Cutoff*-Radius zur Berücksichtigung der Umgebung ein Wert von 2.5 Å gewählt.

- SHI: Zur Berechnung des *Shapeindex* werden zunächst die beiden kanonischen Krümmungen mit dem Programm *globcurv* [38] (*Cutoff* 3.5 Å) bestimmt. Anschließend wird der *Shapeindex* durch das Programm *topoqual* [38] wie in Abschnitt 2.5.5.1 beschrieben berechnet.
- CUR: Die Berechnung der *Curvedness* erfolgt ebenfalls mit Hilfe der mit *globcurv* bestimmten globalen Krümmungen durch das Programm *topoqual*.

4.1.3 Grundlage der statistischen Auswertung

Der Datensatz an Proteinkomplexen wird wie in Abschnitt 3.1 beschrieben in drei Komplexklassen eingeteilt (Antikörper, Dimere, Enzyme). In jeder Komplexklasse werden alle Oberflächenanteile, die Protein-Protein-Kontakte beschreiben, zusammengeführt. Das gleiche wird für die Protein-Kavität-Oberflächenanteile durchgeführt. Die Anzahl der physikochemischen Daten jeder Eigenschaft entspricht somit der Summe aller Oberflächenpunkte, die der jeweiligen Kontaktregion zugeordnet sind. Damit ergibt sich für jede physikochemische Eigenschaft der folgende Umfang an Datenpunkten für die unterschiedlichen Kontaktregionen (Antikörper/Dimere/Enzyme): Protein-Protein-Kontaktregionen (224239/122403/496121) und Protein-Kavität-Regionen (128556/61937/212893). Diese Datensätze stellen die Grundlage für die statistischen Analysen dar. Prozentangaben von Ergebnissen beziehen sich auf die Punktzahl der Teiloberflächen und nicht etwa auf ihren Flächeninhalt. Da die Punktzahl allerdings annähernd proportional zum Flächeninhalt ist, ist der Unterschied zwischen einer punktbasierten und flächenbasierten Auswertung nur marginal.

4.1.4 Berechnung von Volumen und Flächeninhalt

Zur Volumenberechnung der mit dem Programm *space* identifizierten Kavitäten wird zunächst mit dem Programm *fume* eine molekulare Oberfläche für jede Kavität erzeugt. Dazu wird jedem Gitterpunkt der Kavität ein van-der-Waals-Radius von 1.3 Å zugeordnet, was dem „Wasserradius“ beim Auffüllen der Kavität entspricht und somit ihre Ausdehnung approximiert. Der zur Erzeugung der Oberfläche verwendete Abrollradius beträgt wiederum 1.3 Å. Das von der Oberfläche eingeschlossene Volumen entspricht dem

Kavitätsvolumen und wird mit dem Programm *sldinfo* (M. Keil, persönliche Mitteilung) bestimmt. Die Berechnung des Flächeninhalts von Oberflächen bzw. Teiloberflächen erfolgt ebenfalls mit dem Programm *sldinfo*. Angaben über die Größe der Binderegion eines Komplexes beziehen sich im folgenden immer auf die Summe der Teiloberflächen beider Komplexpartner.

4.2 Analyse der Kavitäten- und Wasserverteilung in der Binderegion

Die statistische Analyse der Anzahl an berechneten Wassermolekülen und Kavitäten in der Binderegion für die unterschiedlichen Komplexklassen (Antikörper/Dimere/Enzyme) erlaubt es, Unterschiede oder typische Kennwerte der Wasserbesetzung in Proteinkomplexen zu identifizieren. Gemeinsam ist allen Komplexklassen, daß die Anzahl berechneter Wassermoleküle mit der Größe der Binderegion zunimmt (Abb. 4.2). Der Korrelationskoeffizient liegt bei $r = 0.9$. Im Vergleich dazu ist die Korrelation zwischen experimentell gefundenen Kristallwassermolekülen der Binderegion und der Bindefläche bei einer von Lo Conte *et al.* [17] durchgeführten Studie von 36 Kristallstrukturen mit einem Korrelationskoeffizienten von nur 0.5 deutlich geringer und weniger ausgeprägt.

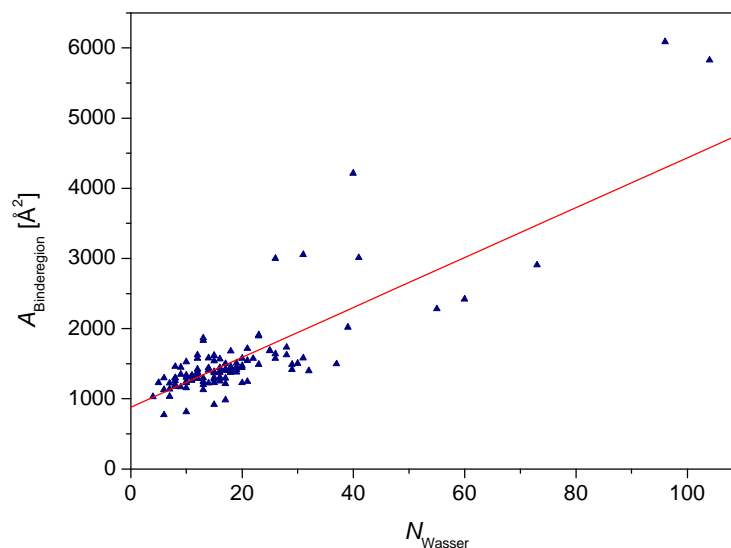


Abb. 4.2: Zusammenhang der berechneten Anzahl an Wassermolekülen N_{Wasser} eines Komplexes mit der Fläche seiner Binderegion $A_{\text{Binderegion}}$. Die lineare Regression (rot) liefert einen Korrelationskoeffizienten von $r = 0.9$. Der Komplex 1bgx wurde wegen seiner großen Anzahl an Wassermolekülen ($N_{\text{Wasser}} = 210$, $A_{\text{Binderegion}} = 7429 \text{ Å}^2$) aus Gründen der Übersichtlichkeit nicht mit in den Graphen aufgenommen.

Zur genaueren Untersuchung der Proteinkomplexe ist die Oberfläche der Binderegion jedes Komplexes wie in Abschnitt 4.1.1 erläutert in unterschiedliche Kontaktregionen eingeteilt. Regionen, die direkte Protein-Protein-Kontakte (P-P-Kontakte) ausbilden, werden von Regionen unterschieden, die eine Kavität ausbilden und daher mit Wassermolekülen der Kavität in Kontakt stehen (K-P-Kontakte). In Abb. 4.3 ist die Verteilung, die den Anteil von Protein-Protein-Kontaktregionen an der gesamten Binderegion angibt, für jeden der drei Datensätze dargestellt.

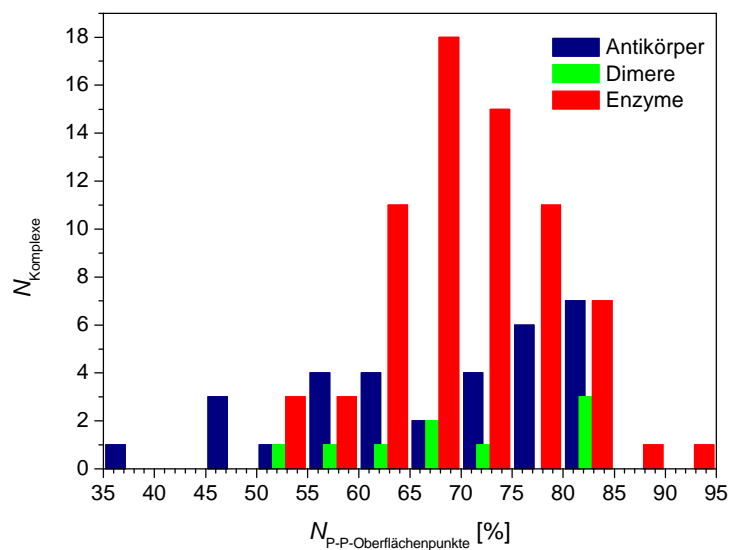


Abb. 4.3: Histogramm der Zusammensetzung aller Binderegionen hinsichtlich ihrer Protein-Protein-Kontaktfläche ($N_{Komplexe}$: Anzahl der Komplexe; $N_{P-P-Oberflächenpunkte}$: Anzahl der Oberflächenpunkte einer Komplexbinderegion, die Protein-Protein-Kontaktregionen darstellen).

Bei Dimeren und Enzymen stellen die Protein-Protein-Kontaktregionen im Mittel 70% der gesamten Binderegion dar, bei Antikörpern liegt der Wert mit 69% etwas darunter. Es fällt auf, daß bei Enzymkomplexen Konfigurationen vorkommen, bei denen die Protein-Protein-Kontaktregionen über 85% der gesamten Binderegion darstellen können. Antikörper weisen hingegen Komplexe auf, bei denen diese Regionen sogar weniger als 50% betragen können. Ein Extremfall ist hierbei der Antikörperkomplex 1bgx mit nur ca. 35% Protein-Protein-Kontaktregionen. Bei diesem Komplex handelt es sich um die TAQ-DNA-Polymerase, an deren DNA-Binderegion ein Antikörperfragment als Inhibitor bindet. Da diese Art von Komplex eher untypisch ist, und die Größe der Binderegion sowie die Anzahl an Wassermolekülen in der Binderegion vergleichsweise hohe Werte aufweisen, wird der Eintrag bei den folgenden Berechnungen von Mittelwerten nicht berücksichtigt.

Neben der unterschiedlichen Verteilung der Kontaktflächen ist auch ein Unterschied bei der Verteilung der Wasseranzahl in den Binderegionen der verschiedenen Proteinkomplexe zu sehen (Abb. 4.4-A). Die mittlere Wasseranzahl pro Komplex beträgt bei Enzymen 17, bei Antikörpern 18 Wassermoleküle. Diese Werte stimmen gut mit Studien experimentell ermittelter Wassermoleküle in hochaufgelösten Kristallstrukturen überein, in denen eine mittlere Anzahl von Wassermolekülen in der Binderegion zwischen 18 und 20 ermittelt wurde [17,26]. Der mittlere Wasseranteil der Dimere liegt mit 45 Molekülen deutlich über den anderen Werten. So zeigt der überwiegende Teil von Enzym- und Antikörperkomplexen eine Wasseranzahl zwischen 2 und 30. Größere Werte treten nur vereinzelt auf; einen Extremfall bildet hierbei wiederum der Antikörperkomplex 1bgx mit 210 Wassermolekülen in der Binderegion. Dimere zeigen demgegenüber einen größeren Anteil an Komplexen mit mehr als 30 Wassermolekülen in der Binderegion. Dies hängt vermutlich damit zusammen, daß 5 der 9 Dimere sehr große Binderegionen von über 2000 \AA^2 besitzen. Der gesamte Datensatz enthält nur 5 weitere Enzym- bzw. Antikörperkomplexe, die ebenfalls eine Bindefläche von mehr als 2000 \AA^2 aufweisen.

Die mittlere Anzahl an Kavitäten in der Komplexbinderegion unterscheidet sich für Antikörper (7 Kavitäten) und Enzyme (6 Kavitäten) kaum. Ihre Häufigkeitsverteilungen bezüglich der Kavitätenanzahl (Abb. 4.4-B) sind sehr ähnlich und weisen Werte zwischen 2 und 13 Kavitäten pro Komplex auf (Ausnahme: Antikörperkomplex 1bgx mit 22 Kavitäten). Im Gegensatz hierzu liegen bei den Dimeren auch Komplexe mit mehr als 13 Kavitäten vor, der Mittelwert von 9 Kavitäten pro Komplex ist ebenfalls erhöht. Die Verteilungen der Wasserbesetzung pro Kavität (Abb. 4.5-A) ähneln sich bei allen drei Komplexklassen. Der Großteil der Kavitäten ist mit 1-3 Wassermolekülen besetzt, es kommen vereinzelt jedoch auch sehr große Kavitäten mit bis zu 55 Wassermolekülen (1bgx) vor. Die mittlere Besetzung liegt für Antikörper und Enzyme bei 3, für Dimere bei 5 Wassermolekülen pro Kavität. Die Verteilung des Kavitätsvolumens (Abb. 4.5-B) stimmt mit dem der Wasseranzahl sehr gut überein, mit zunehmender Wasseranzahl nimmt auch das Kavitätsvolumen zu. Die Volumenverteilung ähnelt vom Verlauf her sehr stark der Häufigkeitsverteilung des Volumens von Kavitäten im Innern globulärer Proteine [58,59]. Die Volumina solcher inneren Kavitäten sind jedoch mit maximal 216 \AA^3 [59] erheblich kleiner als die in der Binderegion identifizierten Kavitäten mit bis zu 1455 \AA^3 . Alle in diesem Abschnitt besprochenen Ergebnisse sind in Tabelle 4.1 zusammengefaßt.

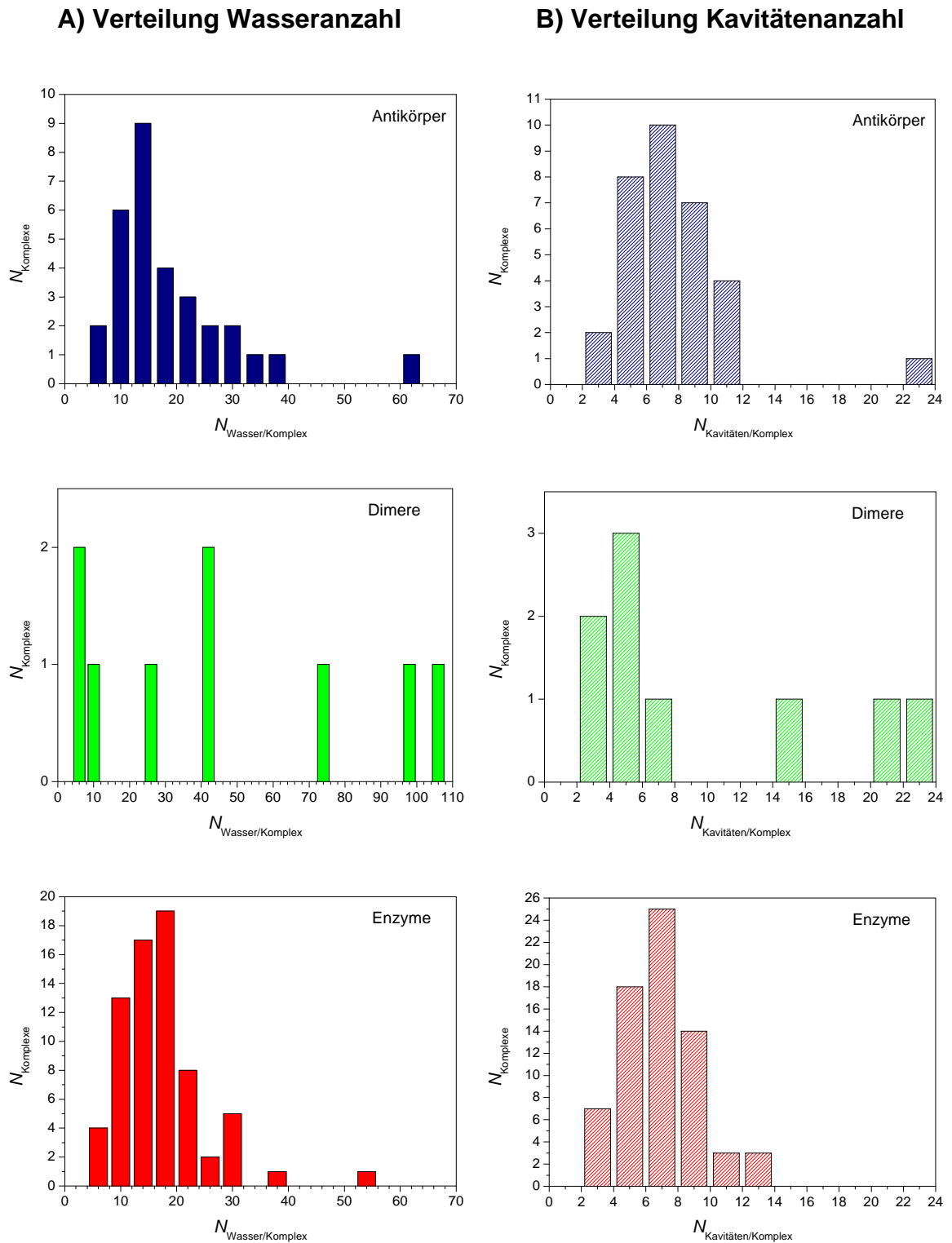


Abb. 4.4: (A) Histogramme der Anzahl an Komplexen N_{Komplexe} in Abhängigkeit von der Anzahl berechneter Wassermoleküle der Binderegion eines Komplexes $N_{\text{Wasser/Komplex}}$. Der Komplex 1bgx (Antikörper) wurde wegen seiner großen Anzahl an Wassermolekülen ($N_{\text{Wasser/Komplex}} = 210$) aus Gründen der Übersichtlichkeit nicht mit in den Graphen aufgenommen. (B) Histogramme der Anzahl an Komplexen N_{Komplexe} in Abhängigkeit von der Anzahl berechneter Kavitäten in der Binderegion eines Komplexes $N_{\text{Kavitäten/Komplex}}$.

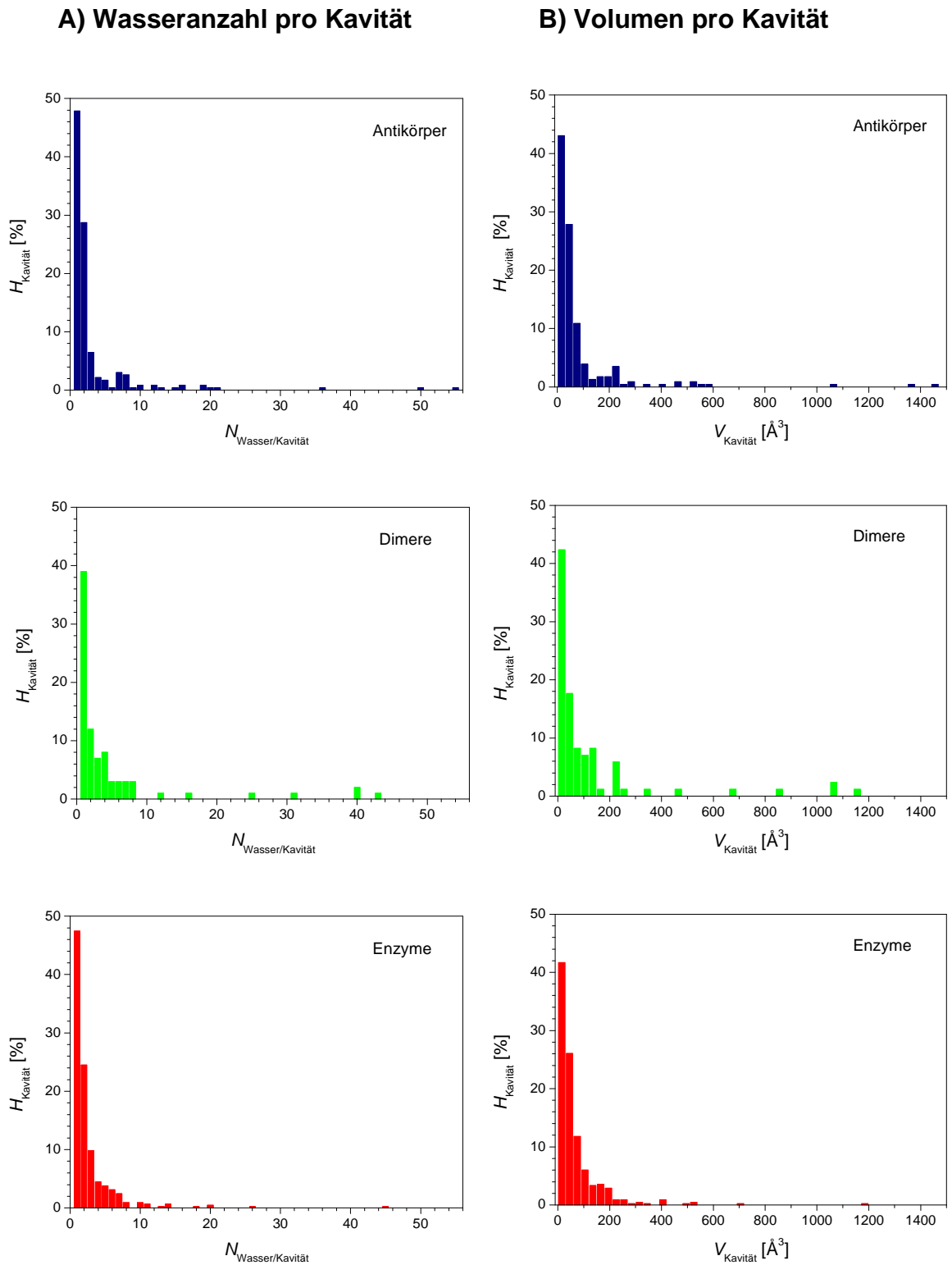


Abb. 4.5: Prozentuale Häufigkeitsverteilungen H_{Komplexe} **A)** der Anzahl von Wassermolekülen innerhalb einer Kavität $N_{\text{Wasser/Kavität}}$ und **B)** des Volumens der Kavitäten $V_{\text{Kavität}}$.

Mittel pro Komplex	Antikörper [†]	Dimere	Enzyme
$\bar{N}_{H_2O} \pm SD$ (Bereich)	17.8 \pm 11.1 (7-60)	44.6 \pm 38.0 (6-104)	16.9 \pm 8.1 (4-55)
$\bar{N}_{Kav} \pm SD$ (Bereich)	6.7 \pm 2.2 (2-22)	9.4 \pm 8.0 (2-23)	6.4 \pm 2.3 (2-13)
$\bar{N}_{H_2O/Kav} \pm SD$ (Bereich)	2.7 \pm 4.4 (1-55)	4.7 \pm 8.4 (1-43)	2.6 \pm 3.5 (1-45)
$\bar{N}_{PP-Punkte} \pm SD$ [%]	68.9 \pm 11.6	70.3 \pm 10.9	70.2 \pm 8.5

Tab. 4.1: Mittelwerte und Standardabweichungen (*SD*) der Wasseranzahl \bar{N}_{H_2O} , der Anzahl an Kavitäten \bar{N}_{Kav} , der Wasseranzahl pro Kavität $\bar{N}_{H_2O/Kav}$ sowie des prozentualen Anteils an Oberflächenpunkten der Binderegion mit Protein-Protein-Kontakt $\bar{N}_{PP-Punkte}$. [†]Berechnung erfolgt ohne Antikörperkomplex 1bgx.

Die Kenntnis der Anzahl von Wassermolekülen in einer Kavität sowie dem Kavitätsvolumen erlaubt es, das Volumen zu berechnen, das einem Wassermolekül in seiner jeweiligen Umgebung zur Verfügung steht. Dabei wird nicht über alle Kavitäten gemittelt, sondern nur über Kavitäten mit gleicher Wasseranzahl, so daß das Volumen eines Wassermoleküls in Abhängigkeit der in der Kavität enthaltenen Wassermoleküle bestimmt werden kann. Diese Volumina sollen mit den aus einer reinen Wasserphase ermittelten Volumina verglichen werden. Zur Bestimmung der Volumina einer reinen Wasserphase wird eine durch Molekulardynamiksimulation erzeugte Wasserbox aus 2500 Wassermolekülen verwendet. Die Equilibrierung der Wasserbox erfolgte bei 298 K und 1 bar in einem *NpT*-Ensemble unter Verwendung des TIP3P-Wassermodells (B. Schilling, persönliche Mitteilung). Aus dieser Wasserbox werden mit dem Programm *randompdb* eine definierte Anzahl an Wassermolekülen zufällig ausgewählt, von denen jedes die Bedingung erfüllen muß, daß das Sauerstoffatom nicht weiter als 3.5 Å von mindestens einem anderen Sauerstoffatom der Auswahl entfernt ist. Für jede Anzahl an Wassermolekülen werden je 100 solcher zufälligen Konfigurationen zusammenhängender Wassermoleküle erzeugt und so das mittlere Volumen pro Wassermolekül in Abhängigkeit der ausgewählten Wasseranzahl bestimmt. Zur Volumenberechnung wird zunächst mit dem Programm *fumee* eine Connolly-Oberfläche um die zusammenhängenden Wassermoleküle generiert und deren Volumen anschließend mit dem Programm *sldinfo* bestimmt. Die Ergebnisse sind in Abb. 4.6 aufgetragen.

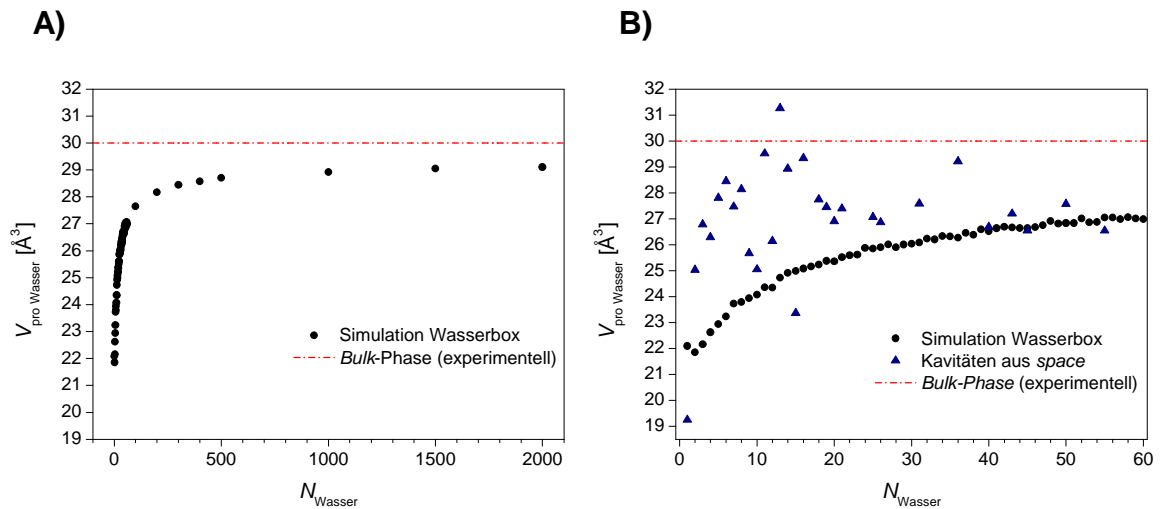


Abb. 4.6: **A)** Mittleres zugängliches Volumen pro Wassermolekül $V_{\text{pro Wasser}}$ (Volumen / Wasseranzahl) in der simulierten Wasserbox in Abhängigkeit von der betrachteten Wasseranzahl N_{Wasser} . Das mittlere Volumen nähert sich dem experimentell ermittelten Wert von Bulk-Wasser (30 \AA^3) mit steigender Wasseranzahl. **B)** Mittleres Volumen pro Wassermolekül $V_{\text{pro Wasser}}$ (Volumen / Wasseranzahl) in den Kavitäten und in der simulierten Wasserbox in Abhängigkeit von der betrachteten Wasseranzahl N_{Wasser} .

Das Volumen eines Wassermoleküls nimmt bei der simulierten Wasserbox mit steigender Wasseranzahl zu und nähert sich bei sehr vielen Wassermolekülen ungefähr dem experimentellem Wert von 30 \AA^3 (entsprechend einer Wasserdichte von 1 kg/l). Die Abweichung beträgt bei 2000 Wassermolekülen ca. 3% zum experimentellen Wert. Eine exakte Übereinstimmung hängt stark von den verwendeten van-der-Waals-Radien sowie dem Radius der Abrollkugel bei der Oberflächenerzeugung ab. Das Volumen der Wassermoleküle in den Kavitäten ist im Vergleich zu dem der Moleküle der Wasserbox größer und zeigt keinen klaren Zusammenhang zwischen Volumengröße und Wasseranzahl. Die Wassermoleküle der Kavitäten scheinen daher im Mittel weniger dicht gepackt zu sein als in der Wasserbox. Im Gegensatz dazu ist bei Kavitäten mit nur einem Wassermolekül zu beobachten, daß diese Wassermoleküle sehr eng zwischen den Komplexpartnern gepackt sein können, da sie ein geringeres Volumen aufweisen als in der Wasserbox. Für eine größere Anzahl an Wassermolekülen (ab 40) stimmen die Volumina der Wasserboxmoleküle und der Wassermoleküle in den Kavitäten jedoch gut überein.

4.3 Analyse der physikochemischen Eigenschaften

Zur Analyse der beiden unterschiedlichen Kontaktbereiche der Binderegion (Protein-Protein und Kavität-Protein) werden die physikochemischen Daten jeder Kontaktregion in Histogrammen gegenübergestellt. Das Histogramm einer berechneten Eigenschaft (z.B. des elektrostatischen Potentials) beschreibt die Häufigkeitsverteilung der Werte, die diese Eigenschaft in dem verwendeten Datensatz annimmt. Im folgenden werden die Histogramme aller berechneten Eigenschaften jedes der drei Datensätze (Antikörper, Dimere, Enzyme) auf Unterschiede hin untersucht.

4.3.1 Elektrostatisches Potential (ESP)

Die Verteilung des elektrostatischen Potentials ähnelt in allen drei Komplexklassen einer Glockenkurve mit dem Maximum am Nullpunkt (Abb. 4.7-A). Enzyme zeigen einen deutlichen Unterschied zwischen Protein-Protein- und Kavitätsregionen. Während bei Werten des elektrostatischen Potentials um den Nullpunkt Protein-Protein-Kontaktregionen den größeren Anteil darstellen, überwiegen bei größerem Potential die Kavitätsregionen. Werte sehr hohen Potentials weisen aber wiederum hauptsächlich Protein-Protein-Kontaktregionen auf, was vermutlich auf die Ausbildung von Salzbrücken zwischen den Komplexpartnern zurückzuführen ist. Ein komplementäres Bild stellt sich bei den Antikörpern dar. Hier überwiegen bei Werten um den Nullpunkt die Kavitätsregionen, während bei stärkeren Potentialwerten die Protein-Protein-Regionen häufiger vertreten sind. Bei Antikörperkomplexen scheinen, abgesehen von spezifischen Salzbrücken, die elektrostatischen Wechselwirkungen direkter Protein-Protein-Kontakte stärker ausgeprägt zu sein als bei Enzymkomplexen. Dimere zeigen ein asymmetrisches Muster bei der Verteilung. Protein-Protein-Kontakte überwiegen bei schwachem Potential. Bei Zunahme in Richtung positiven Potentials überwiegen wie bei den Antikörpern die Protein-Protein-Kontaktregionen, während bei Vergrößerung des Potentials in Richtung negativer Werte wie bei den Enzymen die Kavitätsregionen häufiger vorkommen. Bei den Dimeren ist allerdings zu beachten, daß die statistische Aussagekraft wegen des geringen Datensatzes von nur 9 Komplexen nicht zu hoch bewertet werden darf. Es wird jedoch deutlich, daß sich sowohl die Protein-Protein- von den Kavitätsregionen wie auch die unterschiedlichen Komplexklassen untereinander hinsichtlich der Verteilung des elektrostatischen Potentials unterscheiden.

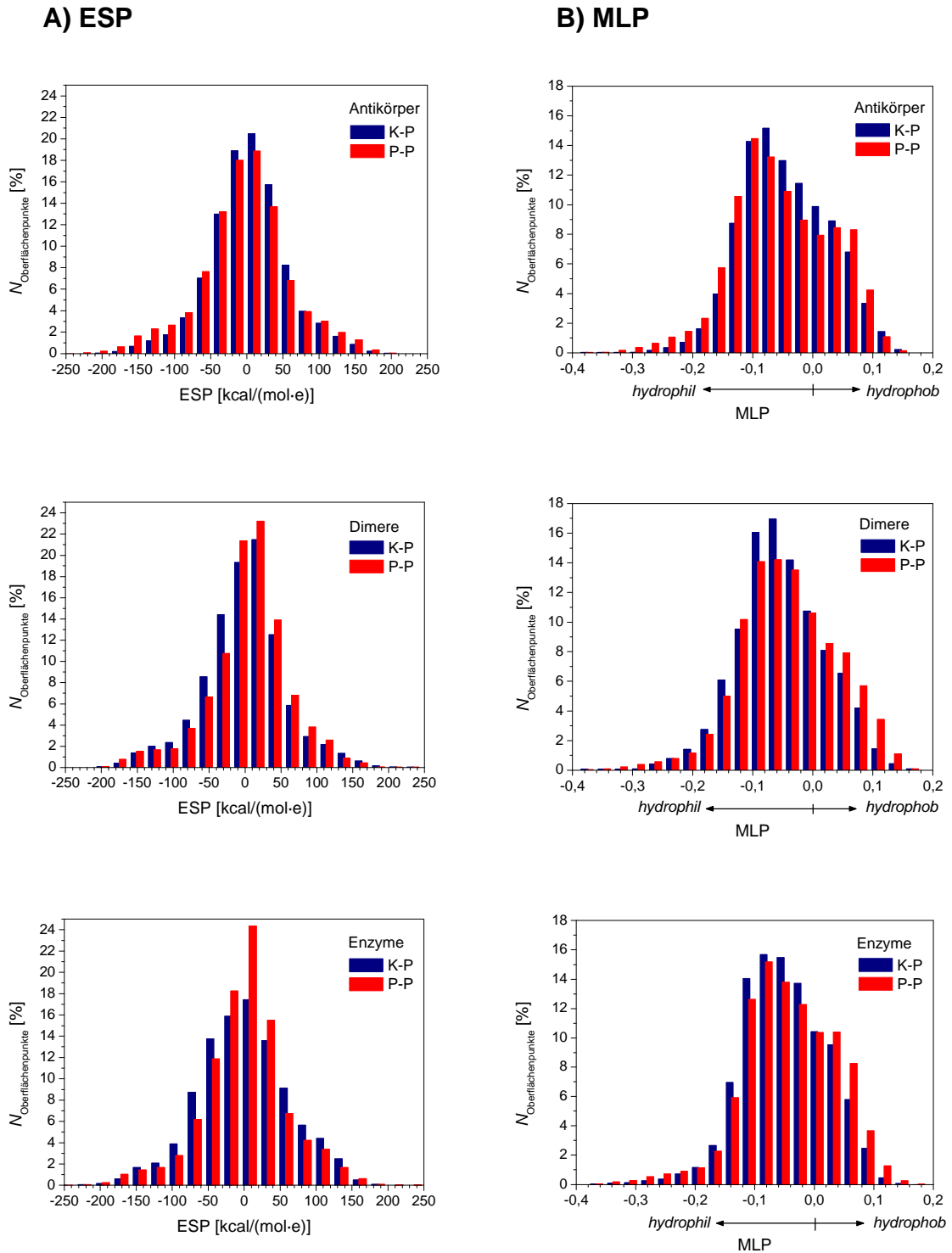


Abb. 4.7: Histogramme der Häufigkeitsverteilung (prozentuale Anzahl Oberflächenpunkte $N_{\text{Oberflächenpunkte}}$) (A) des elektrostatischen Potentials (ESP) und (B) der lokalen Lipophilie (MLP) von Kavität-Protein- (K-P) und Protein-Protein-Kontaktregionen (P-P).

4.3.2 Lokale Lipophilie (MLP)

Die Verteilung der molekularen lokalen Lipophilie (Abb. 4.7-B) verläuft im Gegensatz zum elektrostatischen Potential nicht symmetrisch. Das Maximum der Protein-Protein- wie auch der Kavitätsregionen liegt bei allen drei Komplexklassen bei Werten zwischen -0.06 und -0.08, ist also in den schwach hydrophilen Bereich verschoben. Für hydrophilere (negativere) Werte des Lipophiliepotentials nimmt die Häufigkeit sehr schnell ab, während die Verteilung bis zu schwach hydrophoben Werten (bis +0.06) zunächst langsam, für stärker hydrophobe Werte dann aber schnell abfällt. Für den hydrophoben Bereich zeigen Enzyme und Dimere eine gute Übereinstimmung, die Protein-Protein-Kontaktregionen überwiegen hier gegenüber den Kavitätsregionen. Bei Antikörpern ist dieser Trend nicht über den gesamten hydrophoben Bereich zu beobachten. Im hydrophilen Ast der Kurve überwiegen bei den Enzymen und Dimeren die Kavitätsregionen, mit Ausnahme sehr stark hydrophiler Werte, was vermutlich auf stark polare Wasserstoffbrückenbindungen oder Salzbrücken zurückgeführt werden kann. Antikörper zeigen ein anderes Bild, hier überwiegen die Kavitätsregionen nur bei schwach hydrophilen Werten; bei Werten kleiner als -0.1 weist die Verteilung der Protein-Protein-Kontakte größere Häufigkeiten auf.

4.3.3 Wasserstoffbrücken-Donoren/Akzeptoren (HB2) und ihre Dichte (HD)

Die Häufigkeitsverteilungen von Wasserstoffbrücken-Donoren, -Akzeptoren oder Atomen ohne Wasserstoffbrückenbindungsfähigkeit (HD2) sind für alle drei Komplexklassen sehr ähnlich und lassen keine eindeutigen Unterschiede erkennen (Abb. 4.8-A). Oberflächenanteile ohne Donoren bzw. Akzeptoren stellen sowohl bei Protein-Protein- als auch bei den Kavitätsregionen den größten Anteil dar, wobei die Protein-Protein-Kontaktregionen eine etwas erhöhte Häufigkeit aufweisen. Akzeptoren kommen mit ca. 20% in beiden unterschiedlichen Kontaktregionen der Bindestelle etwas häufiger vor als Donoren. Sowohl bei den Donoren als auch bei den Akzeptoren ist die Häufigkeit der Kavitätsregionen im Vergleich zu den Protein-Protein-Kontaktregionen etwas erhöht. Im Fall der Antikörper ist bei den Akzeptoren praktisch kein Unterschied der Häufigkeiten der unterschiedlichen Kontaktregionen zu erkennen.

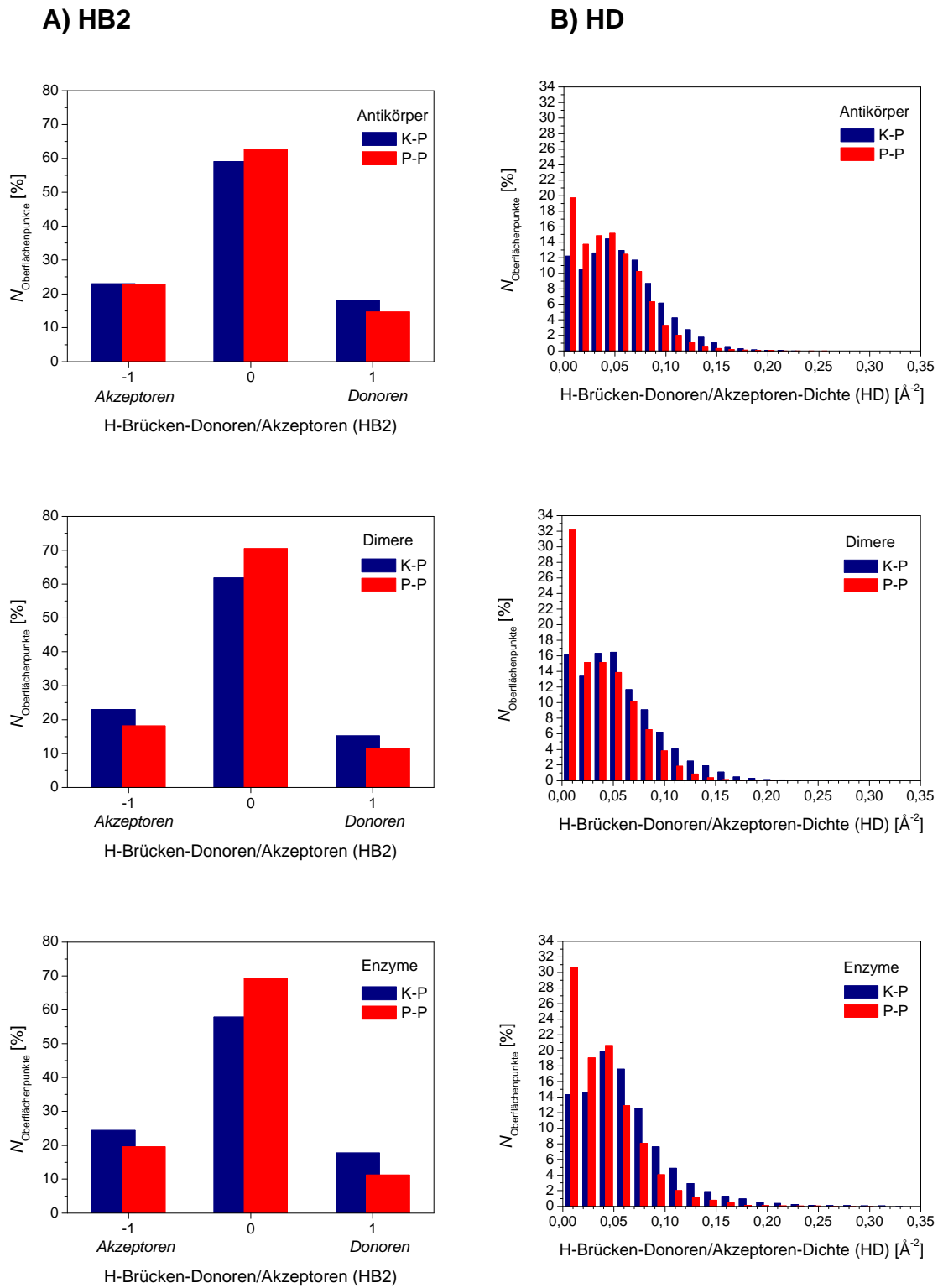


Abb. 4.8: Histogramme der Häufigkeitsverteilung (prozentuale Anzahl Oberflächenpunkte $N_{\text{Oberflächenpunkte}}$) (A) der Wasserstoffbrücken-Donoren bzw. -Akzeptoren (HB2) und (B) der Wasserstoffbrücken-Donoren- bzw. -Akzeptoren-Dichte (HD) von Kavität-Protein- (K-P) und Protein-Protein-Kontaktregionen (P-P).

Auch bei der Donoren-/Akzeptorendichte (HD) sind die Verteilungsfunktionen für die drei Komplexklassen sehr ähnlich (Abb. 4.8-B). Bei sehr geringer Dichte überwiegen eindeutig die Protein-Protein-Kontaktregionen gegenüber den Kavitätsregionen. Im Fall der Antikörper sind die Protein-Protein-Regionen bei sehr geringer Dichte allerdings weniger stark vertreten als bei Enzymen und Dimeren, der Unterschied der beiden Kontaktregionen ist weniger stark ausgeprägt. Mit zunehmender Dichte überwiegen bei allen drei Komplexklassen die Kavitätsregionen gegenüber den Protein-Protein-Kontaktregionen. Es fällt auf, daß sehr hohe Werte der Akzeptoren/Donoren-Dichte nur bei Dimeren und Enzymen erreicht werden, jedoch nicht bei Antikörpern. Bei den Bereichen hoher Dichte handelt es sich ausschließlich um Kavitätsregionen. Möglicherweise sind diese Bereiche notwendig, um konservierte Wassermoleküle optimal an die Proteine zu binden.

4.3.4 Salzbrückenbildner (SB2)

Bei der Verteilung potentieller Salzbrückenbildner in der Binderegion ist bei allen drei Komplexklassen klar zu erkennen, daß über 85% der Oberflächenbereiche beider Kontaktregionen von ungeladenen Atomen gebildet werden (Abb. 4.9-A). Bereiche mit positiv geladenen Nachbargruppen kommen geringfügig häufiger vor als Bereiche mit negativ geladenen Nachbargruppen. Eine Unterscheidung der beiden Kontaktregionen ist für Dimere praktisch nicht möglich, die Unterschiede sind marginal. Bei Antikörpern ist nur ein Unterschied bei den negativ geladenen Gruppen zu erkennen, deren Oberflächenbereiche bei Protein-Protein-Kontakten geringfügig häufiger vorkommen. Bei Enzymen sind die Unterschiede etwas deutlicher zu erkennen. So kommen bei ungeladenen Oberflächenbereichen Protein-Protein-Kontaktbereiche häufiger vor, während bei geladenen Bereichen die Kavitätsregionen eine etwas größere Häufigkeit aufweisen.

4.3.5 Tiefeninformation (CAV)

Die Verteilung der Information über Taschen und Spalten verläuft bei allen drei Komplexklassen praktisch gleich (Abb. 4.9-B). Für beide Kontaktregionen liegen die Maxima bei geringer Taschentiefe, und mit zunehmender Tiefe fallen die Häufigkeiten schnell ab. Protein-Protein-Kontaktregionen weisen bei geringen Taschentiefen sehr viel größere Häufigkeiten auf als Kavitätsregionen. Die Verteilung fällt dann allerdings sehr schnell ab, so daß ab einer Taschentiefe von ca. 1.5 Å die Kavitätsregionen häufiger vorkommen, deren Wassermoleküle gut in den Taschen untergebracht werden können.

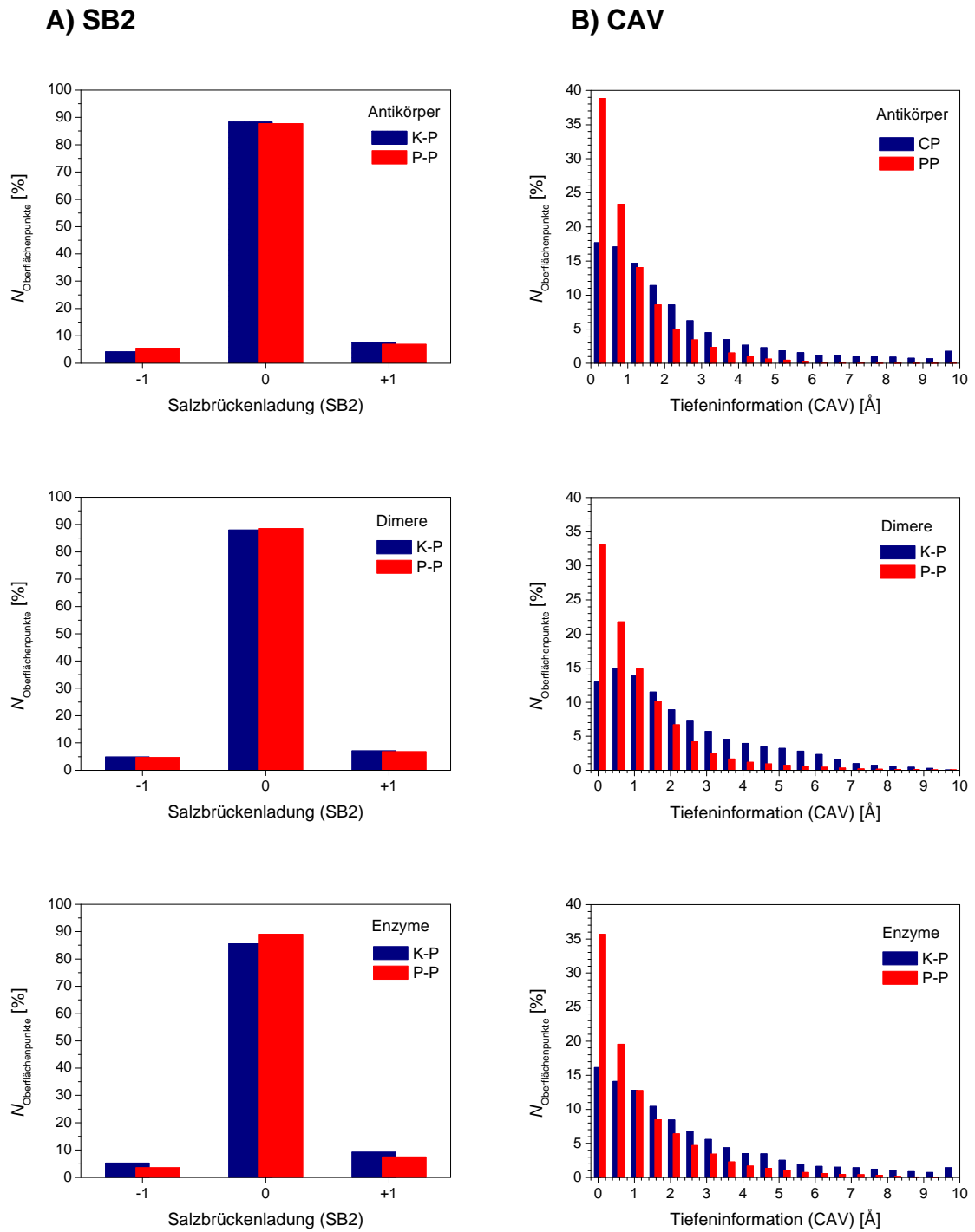


Abb. 4.9: Histogramme der Häufigkeitsverteilung (prozentuale Anzahl Oberflächenpunkte $N_{\text{Oberflächenpunkte}}$) (A) der potentiellen Salzbrückenbildner (SB2) und (B) der Tiefe von Taschen (CAV) von Kavität-Protein- (K-P) und Protein-Protein-Kontaktregionen (P-P).

4.3.6 Shapeindex (SHI) und Curvedness (CUR)

Die Unterschiede zwischen Protein-Protein-Kontaktregionen und Kavitätsregionen sind bei keiner anderen Eigenschaft so deutlich wie bei der Verteilung der Werte des *Shapeindex* (SHI), die für Antikörper, Dimere und Enzyme nahezu identisch verlaufen (Abb. 4.10-A). Bei Protein-Protein-Kontaktregionen nimmt die Häufigkeit von konkaven zu konvexen Formen hin annähernd linear zu. Das Minimum entspricht der Form eines Loches, das Maximum der eines Pfropfes. Mehr als die Hälfte dieser Oberflächenbereiche bilden daher konvexe Formen, die sich als Grat oder Pfropf beschreiben lassen. Kavitätsregionen weisen den entgegengesetzten Verlauf auf. Das Maximum der Verteilung liegt hier im konkaven Bereich, der in etwa der geometrischen Form eines Spaltes entspricht. Dies stimmt mit der Beobachtung überein, daß Wassermoleküle der Binderegion vorwiegend in Löchern und Spalten zwischen den Komplexpartnern liegen und wird durch die Ergebnisse der Tiefeninformation unterstützt, wo Kavitätsregionen bei tiefen Taschen stärker vertreten sind als Protein-Protein-Kontaktregionen. Die Häufigkeitsverteilung der Kavitätsregionen fällt in Richtung konvexer Regionen ab und erreicht bei der geometrischen Form des Pfropfes ihr Minimum.

Bei der Verteilung der Oberflächenneigung, oder *Curvedness* (CUR), unterscheiden sich die beiden unterschiedlichen Kontaktregionen nicht sehr stark. Alle Verteilungen sind asymmetrisch und zeigen ein Maximum bei einer *Curvedness* nahe dem Wert 0.2, von dem aus die Häufigkeiten zu kleineren Werten der *Curvedness* schnell, zu größeren Werten hin etwas langsamer abfallen. In dem Bereich 0.18 bis 0.42 überwiegen leicht die Häufigkeiten der Protein-Protein-Kontaktbereiche, während extrem große Werte (nahe 0.8) nur von Kavitätsregionen gebildet werden. Eine Besonderheit der Verteilung liegt bei einem Wert von ca. 0.58 vor. Hier ist für beide Kontaktregionen eine erhebliche Erhöhung der Häufigkeit zu beobachten, die den kontinuierlichen Verlauf der Verteilung unterbricht. Dabei weisen Kavitätsregionen eine größere Häufigkeit auf als Protein-Protein-Kontaktregionen. Dies wird besonders bei den Enzymen deutlich, wo die Kavitätsregionen bei dem Wert 0.58 eine Häufigkeit von fast 10% annehmen. Eine *Curvedness* von 0.58 scheint wegen der erhöhten Häufigkeit eine geometrisch bevorzugte Form darzustellen. Charakteristische Formen von Oberflächenfragmenten mit diesen Werten bilden beispielsweise den Boden einer Spalte (Abb. 4.11-A) oder sind beim Aufbau von engen Kanälen beteiligt (Abb. 4.11-B), in denen sehr gut Wassermoleküle untergebracht werden können.

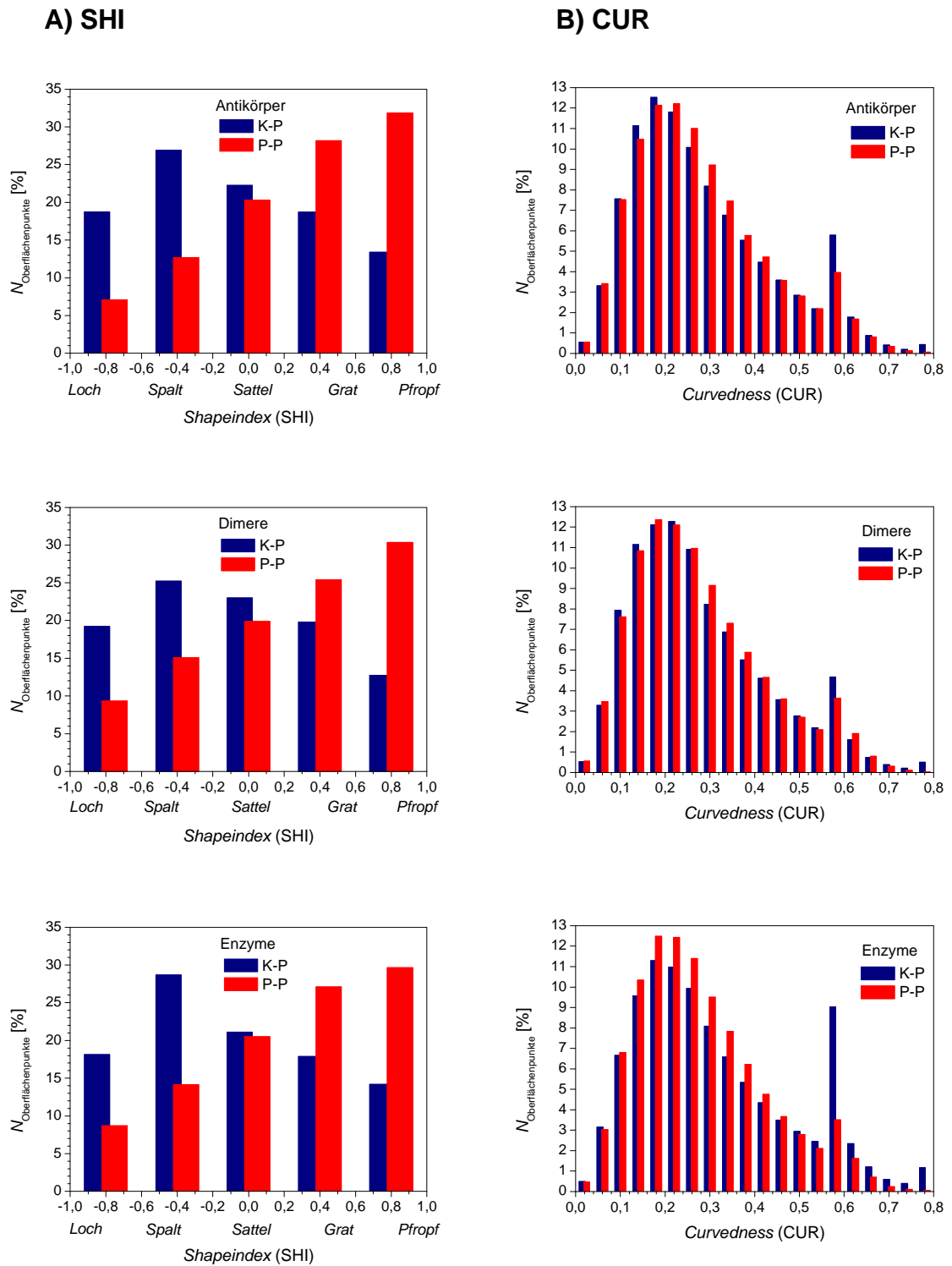


Abb. 4.10: Histogramme der Häufigkeitsverteilung (prozentuale Anzahl Oberflächenpunkte $N_{\text{Oberflächenpunkte}}$) (A) des *Shapeindex* (SHI) und (B) der *Curvedness* (CUR) von Kavität-Protein- (K-P) und Protein-Protein-Kontaktregionen (P-P).

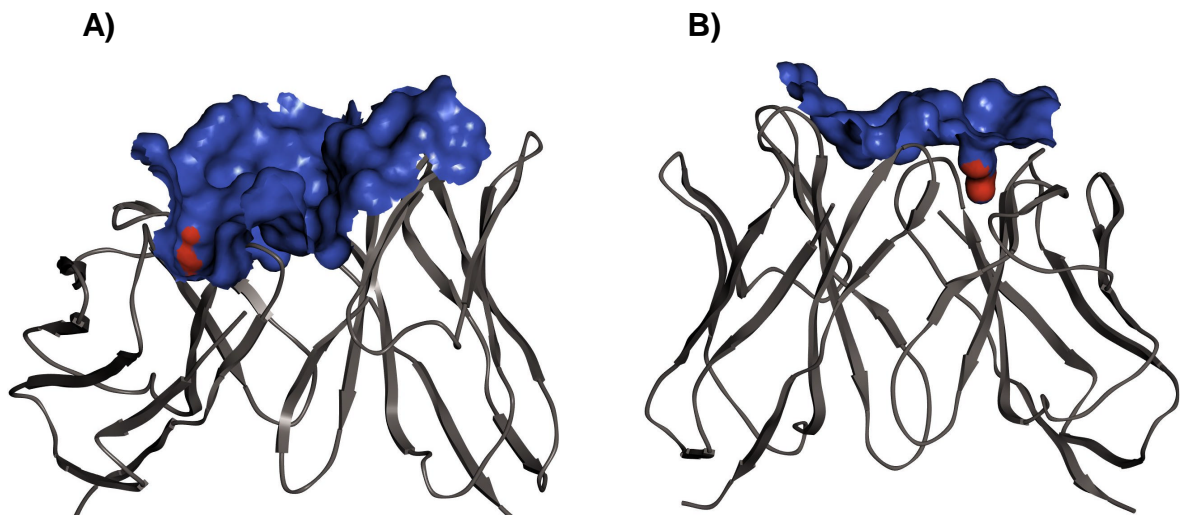


Abb. 4.11: Beispiele zusammenhängender Oberflächenbereiche mit Werten der *Curvedness* zwischen 0.56 und 0.6 (rot). **A)** Binderegion (blau) des Antikörperfragmentes (graues *Ribbon*) aus dem Komplex 3hfl. Die roten Oberflächenbereiche bilden den Boden einer Spalte, die Platz für Wassermoleküle bietet. **B)** Binderegion (blau) des Antikörperfragmentes (graues *Ribbon*) aus dem Komplex 1vfb. Die roten Oberflächenbereiche sind an der Bildung eines kurzen Kanals ins Proteininnere beteiligt, in dem sich Wassermoleküle befinden.

4.4 Zusammenfassung der statistischen Analyseergebnisse

Die statistische Analyse theoretisch ermittelter Wassermoleküle in der Komplexbinderegion zeigt, daß Wassermoleküle bei praktisch allen Komplexen mehr oder weniger stark in der Binderegion vertreten sind. Im Mittel liegen bei Enzymkomplexen 17 und bei Antikörperkomplexen 18 Wassermoleküle in der Binderegion eines Komplexes vor, was sehr gut mit der gemittelten Wasseranzahl aus Untersuchungen mit experimentell ermittelten Kristallwassermolekülen (18-20) [17,26] übereinstimmt. Dimere weisen in der vorliegenden Analyse einen erheblich größeren berechneten mittleren Wasseranteil von 45 Molekülen auf. Für ein gesichertes quantitatives Ergebnis muß der aus nur 9 Strukturen bestehende Datensatz an Dimeren allerdings vergrößert werden. Im Mittel liegen in der Binderegion 7 Kavitäten vor, die durchschnittlich mit 3 Wassermolekülen besetzt sind. Die Anzahl an Kavitäten bzw. Wassermolekülen sowie die Anzahl von Wassermolekülen in den Kavitäten variiert allerdings sehr stark und hängt von dem betrachteten Komplex ab. So existiert beispielsweise ein Komplex, der 210 Wassermoleküle in der Binderegion

aufweist und eine Kavität besitzt, die mit 55 Wassermolekülen besetzt ist. Es ist ein grober Trend zu beobachten, daß mit Zunahme der Binderegionsfläche auch die Wasseranzahl in der Binderegion zunimmt. Das Volumen, das einem Wassermolekül in einer Kavität zur Verfügung steht, ist gegenüber dem Volumen in der reinen *Bulk*-Phase einer mit Molekulardynamik simulierten Wasserbox erhöht. Die Packung der berechneten Wassermoleküle scheint in den Kavitäten nicht ganz so dicht zu sein wie in der Wasserphase, bei größeren Kavitäten stimmen die Volumina mit denen des *Bulk*-Wassers allerdings gut überein.

Mit etwa 70% stellen die Protein-Protein-Kontaktregionen gegenüber den Regionen, die eine Kavität bilden, den größeren Anteil der Bindefläche der Komplexpartner dar. Die beiden unterschiedlichen Kontaktregionen unterscheiden sich vor allem in ihrer Topographie. Kavitätsregionen bilden häufiger konkave Strukturen wie Spalten oder Löcher, in denen sehr gut Wassermoleküle untergebracht werden können, während Protein-Protein-Kontaktregionen bevorzugt konvexe Formen ausbilden. Ein weiterer Unterschied besteht in der leicht erhöhten Wasserstoffbrücken-Donoren/Akzeptoren-Dichte bei Kavitätsregionen, die notwendig ist, um die Wassermoleküle in den Kavitäten mit Wasserstoffbrückenbindungen so gut wie möglich absättigen zu können. Die Protein-Protein-Kontaktregionen überwiegen im Gegensatz dazu bei der Ausbildung hydrophober Oberflächenbereiche. Die beiden Kontaktbereiche unterscheiden sich stärker in ihren geometrischen Eigenschaften, wie *Shapeindex* oder Tiefeninformation, als in ihren physikochemischen Eigenschaften. Die Wasserstoffbrücken-Donoren/Akzeptoren-Dichte weist zwar noch deutliche Unterschiede zwischen den Kontaktregionen auf, die Unterschiede in den restlichen physikochemischen Eigenschaften, wie elektrostatisches Potential oder Lipophilie, sind hingegen wesentlich schwächer ausgeprägt.

Die Aufteilung des Datensatzes in unterschiedliche Komplexklassen erweist sich nach der statistischen Analyse der Wasserverteilung und der physikochemischen Eigenschaften der Binderegion als sinnvoll, da Dimere, Enzym- und Antikörperkomplexe teilweise unterschiedliche Analyseergebnisse aufweisen. Dimere besitzen im Mittel die meisten Wassermoleküle und Kavitäten in der Binderegion, Enzyme die wenigsten. Antikörper unterscheiden sich von Enzymen und Dimeren besonders in den Protein-Protein-Kontaktregionen. Zum einen stellen sie bei den Antikörpern einen etwas geringeren Anteil

der Binderegion dar, zum anderen unterscheiden sie sich in den Eigenschaftsverteilungen nicht so stark von den Kavitätsregionen hinsichtlich Hydrophilie und Polarität. Während bei Enzymen Protein-Protein-Kontaktregionen in hydrophilen Bereichen und bei erhöhtem elektrostatischen Potential weniger häufig vorkommen als Kavitätsregionen, zeigen Antikörper eine umgekehrte Häufigkeitsverteilung der unterschiedlichen Kontaktregionen. Dies deutet darauf hin, daß die Protein-Protein-Kontakte in Antikörpern vermutlich eher durch polare Wechselwirkungen charakterisiert werden können, während diese Kontakte in Enzymen im Vergleich dazu einen größeren Anteil an hydrophoben Wechselwirkungen ausbilden. Diese Beobachtung könnte mit der speziellen physiologischen Funktion der Antikörper zusammenhängen. Es muß gewährleistet sein, daß durch Variation der Aminosäuren in den hypervariablen Regionen der Antikörper sehr schnell eine hohe Spezifität für ein Antigen ausgebildet werden kann. Die Spezifität wird vorwiegend durch elektrostatische, polare Wechselwirkungen erreicht, die bei den Antikörpern im Vergleich zu den Enzymen scheinbar etwas stärker ausgeprägt sind. Natürlich zeigen auch Enzyme eine hohe Spezifität für ihr Substrat, allerdings konnte hierbei die Entwicklung des Schlüssel-Schloß-Prinzips über einen sehr viel größeren Zeitraum erfolgen und dadurch optimiert werden.

5 Vorhersage von Bindungskonstanten

In dem folgenden Kapitel wird die Entwicklung einer Zielfunktion beschrieben, die eine schnelle Abschätzung der Bindungskonstanten von Protein-Protein-Komplexen anhand ihrer strukturellen und physikochemischen Eigenschaften ermöglichen soll. Dabei wird auf ein etabliertes Verfahren zurückgegriffen, dem die Annahme zugrunde liegt, daß sich die logarithmierte Komplexbindungskonstante $\lg(K_A)$ als Summe physikalisch sinnvoller Teilbeträge (Deskriptoren D_i) von strukturellen und physikochemischen Moleküleigenschaften formulieren läßt:

$$\lg(K_A) = C + \sum_i k_i \cdot D_i . \quad (5.1)$$

Die Koeffizienten k_i sowie die Regressionskonstante C der Funktion werden durch multivariate Regression an einen Datensatz von Proteinkomplexen angepaßt, für den die Werte der Deskriptoren berechnet wurden und die experimentellen Bindungskonstanten bekannt sind. Die so erhaltene Funktion wird häufig als *Scoring*-Funktion bezeichnet. Die Regressionskonstante C enthält dabei die Faktoren, die ebenfalls Einfluß auf die Bindeaffinität haben, aber in den verwendeten Deskriptoren nicht berücksichtigt werden. Als Deskriptoren kommen Eigenschaften in Frage, die an den Protein-Protein-Wechselwirkungen beteiligt sind und die Stabilität des Komplexes beeinflussen (Vgl. Abschnitt 2.1 und 2.2). Dazu zählen beispielsweise das elektrostatische Potential, die lokale Lipophilie, Wasserstoffbrückenbindungen, Salzbrücken oder ein Maß für die durch Komplexbildung eingeschränkten Rotationsfreiheitsgrade der Aminosäureseitenketten. Daneben scheint auch die Größe der durch Komplexbildung begrabenen Oberfläche ein guter Deskriptor zur Abschätzung der Bindeaffinität zu sein [33,34]. Für Komplexe zwischen Proteinen und kleinen Liganden existieren inzwischen eine Reihe von *Scoring*-Funktionen, die diese oder ähnliche Moleküleigenschaften als Deskriptoren verwenden [69-72]. Zielfunktionen für Protein-Protein-Komplexe wurden zwar ebenfalls entwickelt, allerdings umfassen die dabei verwendeten Testdatensätze gerade einmal um die 20 Proteinkomplexe [73-75]. Darüber hinaus bleiben Informationen über Kavitäten und Wassermoleküle innerhalb der Binderegion bei all diesen Zielfunktionen unberücksichtigt. Das in der vorliegenden Arbeit entwickelte Verfahren zur Berechnung von Kavitäten und

Wassermolekülen in der Komplexbinderegion ermöglicht es nun recht einfach, solche Informationen als Deskriptoren in eine Zielfunktion zu integrieren. Im folgenden wird untersucht, wie die Wasserinformationen in eine Zielfunktion einfließen können und welche strukturellen und physikochemischen Moleküleigenschaften zusätzlich in der Zielfunktion berücksichtigt werden müssen. Der zur Entwicklung der *Scoring*-Funktion zusammengestellte Datensatz an Proteinkomplexen ist mit 42 PDB-Einträgen umfangreicher als die der oben erwähnten Studien und wird im folgenden Abschnitt näher besprochen.

5.1 Auswahl des Datensatzes

Alle Strukturen des zur Entwicklung der *Scoring*-Funktion verwendeten Datensatzes (Datensatz II) sind der PDB entnommen und stellen ausschließlich Protein-Protein-Komplexe (Antikörper-Antigen bzw. Enzym-Substrat/Inhibitor) dar, deren experimentelle Bindungskonstanten der Literatur entnommen werden können. Wie bei dem unter Abschnitt 3.1 beschriebenen Datensatz werden Proteinketten, die eine funktionelle Einheit darstellen und einen Komplexpartner bilden, zusammengefaßt. Die Binderegion jedes Komplexpartners besteht ausschließlich aus den 20 Standardamino-säuren und beinhaltet weder Metall- noch Heteroatome. Enthält ein PDB-Eintrag mehrere Strukturen des gleichen Komplexes, wird jede Komplexstruktur als eigenständiger Komplex in den Datensatz integriert. So besteht der verwendete Datensatz II aus insgesamt 48 Komplexen, die aus 42 PDB-Einträgen stammen. Eine Zusammenstellung der Komplexe mit ihren experimentellen Bindeaffinitäten ist in Tabelle 5.1 aufgeführt. Die Bindeaffinität der verwendeten Komplexe reicht von 10^5 M^{-1} bei dem Komplex aus Cytochrom c und Cytochrom c Peroxidase bis $2 \cdot 10^{13} \text{ M}^{-1}$ im Falle des Enzym-Inhibitor-Komplexes β -Trypsin/PTI. Bindeaffinitäten, die in der Literatur als Freie Enthalpie angegeben sind, werden über die Beziehung $\Delta_r G^\ominus = -RT \cdot \ln K_A$ mit $T = 298 \text{ K}$ in die entsprechende Bindungskonstante K_A umgerechnet. Die Bearbeitung der Kristallstrukturen erfolgt wie unter Abschnitt 3.2 beschrieben, d.h. fehlende Atome wie Wasserstoffatome werden addiert und durch Energieminimierung relaxiert. Die Kristallwassermoleküle werden für die Entwicklung der *Scoring*-Funktion nicht verwendet und werden daher im folgenden nicht weiter berücksichtigt.

PDB-ID	Komponenten (Ketten-ID)	lg(K_A)	Lit.
1a14	Neuraminidase (N) / Anti-Neuraminidase F _v NC10 (H/L)	7.85	[121]
1a2y	Antibody D1.3 (A/B) / Lysozyme D18A (C)	7.65	[122]
1acb	α -Chymotrypsin (E) / Eglin C (I)	10.70	[123]
1ar1	Cytochrome c Oxidase (A/B) / F _v (C/D)	10.00	[121]
1bql	HyHEL-5 F _{ab} (H/L) / Lysozyme (Y)	6.70	[124]
1ca0_1;2	Bovine Chymotrypsin (A/B/C/F/G/H) / APPI (D;I)	8.24	[125]
1cbw_1;2	Bovine Chymotrypsin (A/B/C/F/G/H) / BPTI (D;I)	7.94	[126]
1cgi	α -Chymotrypsinogen (E) / HPSTI V3 (I)	10.80	[127]
1cgj	α -Chymotrypsinogen (E) / HPSTI V4 (I)	10.62	[127]
1cho	α -Chymotrypsin (E) / OMTKY3 (I)	10.60	[75]
1cse	Subtilisin Carlsberg (E) / Eglin C (I)	9.60	[75]
1dzb_1;2	scF _v 1F9 (A;B) / Turkey Egg-White Lysozyme C (X;Y)	6.28	[128]
1fbi_1;2	F _{ab} F9.13.7 (L/H;P/Q) / Lysozyme (X;Y)	8.00	[129]
1fdl	Anti-Lysozyme F _{ab} D1.3 (L/H) / Lysozyme (Y)	8.43	[46]
1jhl	Anti-HE-Lysozyme F _v D11.15 (L/H) / PE-Lysozyme (A)	10.18	[130]
1kip	Antibody D1.3 Y(B32)A (A/B) / Lysozyme (C)	7.85	[46]
1kiq	Antibody D1.3 Y(B101)F (A/B) / Lysozyme (C)	6.60	[46]
1kir	Antibody D1.3 Y(A50)S (A/B) / Lysozyme (C)	7.41	[46]
1mah	Acetylcholinesterase (A) / Fasciculin 2 (F)	11.64	[131]
1mct	Trypsin (A) / CMTI-I (I)	10.60	[132]
1mlc_1;2	F _{ab} D44.1 (A/B;C/D) / HE-Lysozyme (E;F)	7.00	[133]
1ppe	Trypsin (E) / CMTI-I (I)	11.52	[134]
1ppf	Human Leukocyte Elastase (E) / OMTKY3 (I)	9.90	[75]
1sbn	Subtilisin Novo BPN' (E) / Eglin C L45R (I)	10.27	[135]
1sib	Subtilisin Novo BPN' (E) / Eglin C R53K (I)	9.96	[135]
1taw	Trypsin (A) / APPI (B)	9.96	[125]
1tec	Thermitase (E) / Eglin C (I)	10.27	[75]
1tpa	Anhydro-Trypsin (E) / PTI (I)	13.05	[27]
1vfb	F _v D1.3 (A/B) / HE-Lysozyme (C)	8.36	[27]
2btc	Trypsin (E) / CPTI-II (I)	11.93	[136]
2iff	HyHEL-5 F _{ab} (H/L) / Lysozyme R68K (Y)	8.00	[121]
2kai	Kallikrein A (A/B) / BPTI (I)	9.17	[75]
2pcb	Cytochrome c Peroxidase (A) / Cytochrome c (B)	5.00	[137]
2pcc_1;2	Cytochrome c Peroxidase (A;C) / Iso-1-Cytochrome c (B;D)	5.88	[138]
2ptc	β -Trypsin (E) / PTI (I)	13.27	[75]
2sec	Subtilisin Carlsberg (E) / N-Acetyl-Eglin C (I)	10.27	[75]
2sic	Subtilisin BPN' (E) / Streptomyces Subtilisin Inhibitor (I)	9.31	[75]
2sni	Subtilisin Novo (E) / Chymotrypsin Inhibitor 2 (I)	11.59	[75]
2tgp	Trypsinogen (Z) / PTI (I)	13.05	[75]
3hfl	HyHEL-5 F _{ab} (H/L) / Lysozyme (Y)	10.41	[75]
3sgb	Proteinase B SGPB (E) / OMTKY3 (I)	9.31	[75]
4sgb	Serine Proteinase B (E) / PCI-1 (I)	8.58	[75]

Tab. 5.1: PDB-Einträge mit Proteinennamen und zugehöriger Ketten-ID des Datensatzes II an Protein-Protein-Komplexen zur Entwicklung einer *Scoring*-Funktion. Einträge mit mehreren Strukturen sind durch einen Unterstrich markiert. Experimentelle Bindungskonstanten K_A sind der angegebenen Literatur entnommen.

5.2 Berechnung der Deskriptoren

Für die Entwicklung einer Zielfunktion zur Vorhersage von Komplexbindungskonstanten ist die Auswahl sinnvoller Deskriptoren von großer Bedeutung. Als Deskriptoren kommen beispielsweise das elektrostatische Potential, die lokale Lipophilie, Wasserstoffbrückenbindungen oder Salzbrücken in Frage. Diese Größen können in unterschiedlicher Form als Deskriptoren verwendet werden. Häufig werden polare Kontaktpaare abgezählt oder, wie auch in dieser Arbeit, die Anzahl an Wasserstoffbrücken (D_{nh}) bestimmt. Allerdings wird in der vorliegenden Arbeit für das elektrostatische Potential (D_{esp}), die lokale Lipophilie (D_{mlp}), die Wasserstoffbrückenbindungen (D_{hb2}) und die Salzbrücken (D_{sb2}) erstmals eine neue Vorgehensweise zur Berechnung von Deskriptoren mit Hilfe des Konzepts der *Separating Surface* (Abschnitt 2.4.3) eingeschlagen. Diese Oberfläche eignet sich sehr gut zur Bewertung der Komplementarität der beiden Komplexpartner bezüglich einer betrachteten Eigenschaft, da die Informationen beider Komplexpartner auf die gleichen Raumpunkte projiziert werden. Damit läßt sich für jeden Punkt ein einfaches Bewertungsschema erstellen, wie gut die Komplexpartner hinsichtlich der betrachteten Eigenschaft zusammenpassen. Durch Summation über alle Oberflächenpunkte kann anschließend eine Art Komplementaritätsindex hinsichtlich der betrachteten Eigenschaft berechnet werden, der als Deskriptor für die Zielfunktion verwendet werden kann. Darüber hinaus werden auch herkömmliche Deskriptoren wie die Oberflächengrößen unterschiedlicher Binderegionsbereiche (D_a , D_{PPa} , D_{al} , D_{PPal}) berechnet. Des weiteren werden erstmals Informationen über Volumen und Oberflächengröße von Kavitäten (D_k), Kavitätenanzahl (D_{nk}) sowie über die berechnete Wasseranzahl (D_{nw}) innerhalb der Binderegion als Deskriptoren getestet. In Tabelle 5.2 sind nochmals alle Deskriptoren zusammengestellt.

D_a : gesamte Binderegionsfläche	D_{sb2} : Salzbrücken (S.S.)
D_{al} : lipophile Binderegionsfläche	D_{hb2} : Wasserstoffbrücken (S.S.)
D_{PPa} : Protein-Protein-Kontaktfläche	D_{nh} : Anzahl Wasserstoffbrücken
D_{PPal} : lipophile Protein-Protein-Kontaktfläche	D_k : Formfaktor Kavitäten
D_{esp} : elektrostatisches Potential (S.S.)	D_{nk} : Anzahl Kavitäten
D_{mlp} : molekulare Lipophilie (S.S.)	D_{nw} : Anzahl Wasser in Kavitäten

Tab. 5.2: Zusammenstellung der Deskriptoren zur Entwicklung einer Zielfunktion zur Affinitätsvorhersage. Die Abkürzung (S.S.) kennzeichnet Deskriptoren, die mit Hilfe der *Separating Surface* berechnet werden.

5.2.1 Berechnung der *Separating Surface*

Bevor die Deskriptoren des elektrostatischen Potentials, der Lipophilie, der Salz- und Wasserstoffbrücken berechnet werden können, muß zunächst für jeden Komplex die *Separating Surface* generiert und anschließend die jeweilige Eigenschaft jedes Komplexpartners auf die Oberfläche projiziert werden. Zur Berechnung der *Separating Surface* wird das Programm *grisu2* (Gitterabstand 0.5 Å, Radius 1.3 Å) von M. Keil [95] verwendet. Da die Bewertung der Komplementarität der oben aufgeführten Moleküleigenschaften nur für solche Oberflächenanteile sinnvoll ist, die direkte Protein-Protein-Kontaktregionen beschreiben, müssen alle Teile der *Separating Surface*, die in einer Kavität oder am Außenrand der Binderegion liegen und Wassermolekülen zugänglich sind, entfernt werden. Dies wird gewährleistet, indem alle Teile der *Separating Surface* mit dem Programm *cutsurf* herausgeschnitten werden, die weiter als 1.3 Å von der van-der-Waals-Oberfläche der beiden Komplexpartner entfernt sind, oder deren Abstand zu den Gitterpunkten, die laut *space*-Verfahren eine Kavität beschreiben, weniger als 1.3 Å beträgt. Anschließend wird das elektrostatische Potential (ESP), die lokale Lipophilie (MLP), die Verteilung der Wasserstoffbrücken-Akzeptoren/Donatoren (HB2) sowie die Salzbrückenverteilung (SB2) wie unter Abschnitt 4.1.2 beschrieben für jeden Komplexpartner berechnet und auf die gemeinsame *Separating Surface* projiziert (Abb. 5.1).

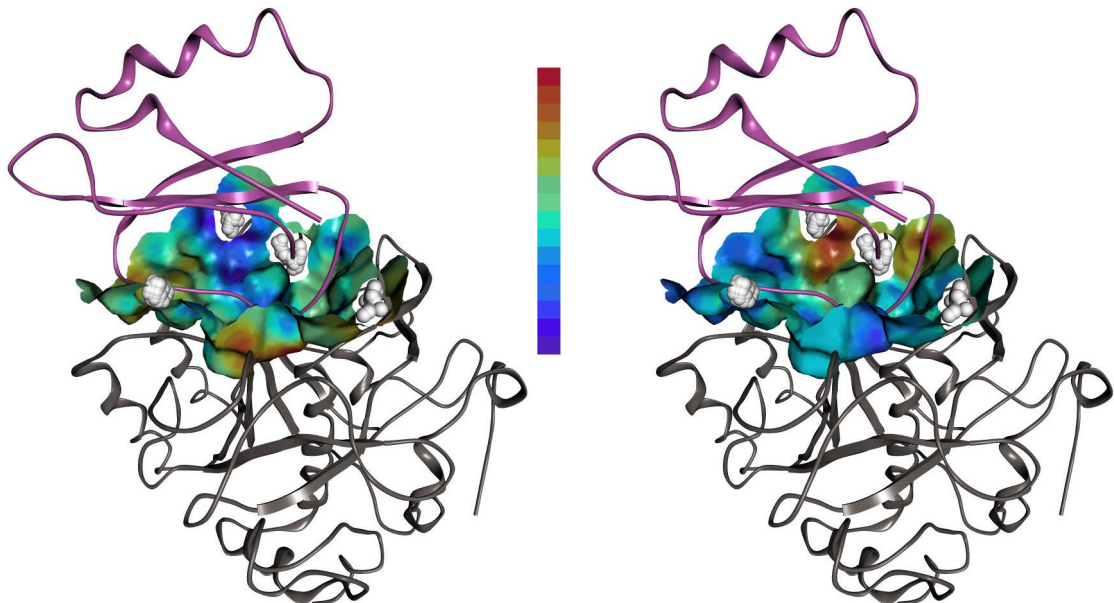


Abb: 5.1: *Separating Surface* des Komplexes 1tec aus dem Enzym Thermitase (grau) und dem Inhibitor Eglin C (magenta). Kavitäten in der Binderegion werden durch weiße Kugeln auf den entsprechenden Gitterpunkten beschrieben. Das elektrostatische Potential wurde für jeden Komplexpartner separat berechnet und ist farbcodiert auf der *Separating Surface* dargestellt (blau: negativ, rot: positiv). Komplementäre Bereiche des elektrostatischen Potentials von **A**) Thermitase und **B**) Eglin C sind gut zu erkennen.

Durch Kombination der unterschiedlichen Werte der beiden Komplexpartner kann nun für jede Moleküleigenschaft ein Deskriptor ermittelt werden. Die genaue Vorgehensweise hierzu wird in den folgenden Abschnitten näher erläutert.

5.2.2 Elektrostatisches Potential D_{esp}

Die Werte des elektrostatischen Potentials der beiden Komplexpartner werden für jeden Oberflächenpunkt i der *Separating Surface* zunächst nach der folgenden Gleichung miteinander kombiniert:

$$d_{\text{esp},i} = \frac{-(\mathbf{j}_{1,i} \cdot \mathbf{j}_{2,i})}{|\mathbf{j}_{1,i}| + |\mathbf{j}_{2,i}|} \quad (5.2)$$

mit $\mathbf{j}_{1,i}$: elektrostatisches Potential von Protein 1 am Oberflächenpunkt i ,
 $\mathbf{j}_{2,i}$: elektrostatisches Potential von Protein 2 am Oberflächenpunkt i .

Durch die Bildung des Quotienten in Gleichung 5.2 wird erreicht, daß große Differenzen der Beträge des elektrostatischen Potentials der Komplexpartner bei gleichem Produkt $\mathbf{j}_{1,i} \cdot \mathbf{j}_{2,i}$ schwächer gewichtet werden als kleine Differenzen der Beträge. Das negative Vorzeichen der Gleichung dient dazu, daß attraktive Wechselwirkungen komplementärer Ladungen positive Werte für $d_{\text{esp},i}$ liefern. Die Summation aller Werte von $d_{\text{esp},i}$ über die gesamte *Separating Surface* liefert den Deskriptor D_{esp} :

$$D_{\text{esp}} = \sum_i d_{\text{esp},i} \quad (5.3)$$

5.2.3 Lokale Lipophilie D_{mlp}

Durch den Deskriptor D_{mlp} sollen hydrophobe Wechselwirkungen berücksichtigt werden. Eine Kombination von Werten der lokalen Lipophilie auf der *Separating Surface* findet daher nur statt, wenn beide Komplexpartner am Oberflächenpunkt i einen lipophilen Bereich aufweisen, also positive Werte besitzen. Andernfalls handelt es sich nicht um hydrophobe Wechselwirkungen, so daß der Wert an diesem Oberflächenpunkt Null gesetzt wird. Die Kombinationsvorschrift lautet also wie folgt:

$$d_{\text{mlp},i} = \begin{cases} 0 \\ \sqrt{MLP_{1,i} \cdot MLP_{2,i}} & \text{für } MLP_{1,i} > 0 \text{ und } MLP_{2,i} > 0 \end{cases} \quad (5.4)$$

mit $MLP_{1,i}$: lokale Lipophilie von Protein 1 am Oberflächenpunkt i ,
 $MLP_{2,i}$: lokale Lipophilie von Protein 2 am Oberflächenpunkt i .

Bei der Kombination der Werte wird das geometrische Mittel gewählt, da das Ergebnis bei stark unterschiedlichen Lipophilie-Werten der Komplexpartner im Vergleich zum arithmetischen Mittel geringer ist, so daß diese Art von Konfiguration schlechter bewertet wird. Der Deskriptor D_{mlp} wird als Summe aller Oberflächenwerte $d_{\text{mlp},i}$ definiert:

$$D_{\text{mlp}} = \sum_i d_{\text{mlp},i} \cdot \quad (5.5)$$

5.2.4 Salzbrücken D_{sb2} und Wasserstoffbrücken $D_{\text{hb2}} / D_{\text{nh}}$

Die Fähigkeit zur Ausbildung von potentiellen Wasserstoffbrücken wird auf den Oberflächenpunkten entweder durch den Wert -1 (H-Akzeptor) oder +1 (H-Donor) ausgedrückt (Kapitel 2.5.3). Ein Wert von Null bedeutet, daß sich kein Wasserstoffakzeptor bzw. -donor in der nahen Umgebung befindet. Analog wird für die Salzbrücken angegeben, ob eine positiv oder negativ geladene Gruppe benachbart ist, die Salzbrücken ausbilden kann, oder ob eine ungeladene Region vorliegt (Kapitel 2.5.4). Zur Ausbildung von Wasserstoff- bzw. Salzbrücken kommt es dann, wenn eine komplementäre Paarung der Werte beider Komplexpartner vorliegt. Dies ist der Fall, wenn das Produkt dieser beiden Werte negativ ist. Die Kombination lautet daher sowohl für Wasserstoffbrücken wie auch für Salzbrücken:

$$d_{\text{hb2},i} = \begin{cases} 0 & \text{für } HB_{1,i} \cdot HB_{2,i} \geq 0 \\ 1 & \text{für } HB_{1,i} \cdot HB_{2,i} < 0 \end{cases} \quad (5.6)$$

mit $HB_{1,i}$: Wasserstoffbrückenbindungsfähigkeit von Protein 1 am Oberflächenpunkt i ,
 $HB_{2,i}$: Wasserstoffbrückenbindungsfähigkeit von Protein 2 am Oberflächenpunkt i ,

$$d_{sb2,i} = \begin{cases} 0 & \text{für } SB_{1,i} \cdot SB_{2,i} \geq 0 \\ 1 & \text{für } SB_{1,i} \cdot SB_{2,i} < 0 \end{cases} \quad (5.7)$$

mit $SB_{1,i}$: Salzbrückenbindungsfähigkeit von Protein 1 am Oberflächenpunkt i ,
 $SB_{2,i}$: Salzbrückenbindungsfähigkeit von Protein 2 am Oberflächenpunkt i .

Der Deskriptor D_{hb2} wird als Summe aller Oberflächenwerte $d_{hb2,i}$ definiert:

$$D_{hb2} = \sum_i d_{hb2,i} \cdot \quad (5.8)$$

Der Deskriptor D_{sb2} wird als Summe aller Oberflächenwerte $d_{sb2,i}$ definiert:

$$D_{sb2} = \sum_i d_{sb2,i} \cdot \quad (5.9)$$

Für die Wasserstoffbrückenbindungsfähigkeit wird neben dem Deskriptor D_{hb2} , der auf Grundlage der *Separating Surface* berechnet wird, ein weiterer, alternativer Ansatz zur Berücksichtigung von Wasserstoffbrücken verfolgt. Der Deskriptor D_{nh} entspricht dabei der Anzahl an Wasserstoffbrückenbindungen, die sich zwischen den beiden Komplexpartnern ausbilden können. Die Anzahl der Wasserstoffbrücken wird mit dem Programm *hbonding* (M. Keil, persönliche Mitteilung) bestimmt. Die Parameter werden dabei so gewählt, daß der Donor-Akzeptor-Abstand maximal 3.2 Å und der Wasserstoff-Akzeptor-Abstand maximal 3.0 Å betragen darf. Außerdem darf der Winkel zwischen Donoratom, Wasserstoff und Akzeptoratom einen Wert von 90° nicht unterschreiten.

5.2.5 Kavitäten- und Wasserinformation D_{nk} , D_{nw} , D_k

Die Anzahl an Kavitäten (D_{nk}) sowie die Anzahl an Wassermolekülen (D_{nw}) in den Kavitäten der Binderegion eines Komplexes werden wie unter Abschnitt 3.3 beschrieben mit dem *space*-Verfahren für jeden Komplex des Datensatzes II berechnet. Anschließend wird das Gesamtvolumen aller Kavitäten eines Komplexes sowie der Flächeninhalt der sie umgebenden Oberflächen wie in Abschnitt 4.1.4 bestimmt. Der Quotient aus Volumen und Oberfläche der Kavitäten stellt den Formfaktor der Kavitäten dar und wird durch den Deskriptor D_k beschrieben.

5.2.6 Größe unterschiedlicher Bindereionsbereiche D_a , D_{PPa} , D_{al} , D_{PPal}

Ausgangspunkt zur Berechnung der unterschiedlichen Bindereionsbereiche sind die molekularen Oberflächen der einzelnen Komplexpartner, die schon bei der Bestimmung der Wasserinformation mit dem *space*-Verfahren generiert wurden. Die Bindereion wird wie unter Abschnitt 4.1.1 erläutert identifiziert, in Protein-Protein- und Protein-Kavit t-Kontaktregionen unterteilt und die jeweiligen Bereiche aus der Gesamtoberfl che ausgeschnitten. Die Berechnung des Oberfl cheninhaltes der unterschiedlichen Regionen wird mit dem Programm *sldinfo* (M. Keil, pers nliche Mitteilung) durchgef hrt und stellt im folgenden die Summe aus beiden Komplexpartnern dar. Der Deskriptor D_a entspricht nun der Fl che der gesamten Bindereion, w hrend der Deskriptor D_{PPa} die Fl che der Protein-Protein-Kontaktbereiche darstellt. Um die Gr  e hydrophober Fl chenanteile der Bindereion bestimmen zu k nnen, wird wie unter Abschnitt 4.1.2 beschrieben die lokale Lipophilie auf den Oberfl chenregionen berechnet. Anschlie end werden mit dem Programm *cutsurf* (M. Keil, pers nliche Mitteilung) alle Oberfl chenbereiche mit Werten kleiner Null ausgeschnitten. Die verbleibenden Oberfl chen stellen ausschlie lich hydrophobe Bereiche dar. Die Gr  e der hydrophoben Oberfl chenbereiche der gesamten Bindereion werden durch den Deskriptor D_{al} repr sentiert; die Fl che der hydrophoben Bereiche der Protein-Protein-Kontaktregionen entspricht dem Deskriptor D_{PPal} .

5.3 Auswahl optimaler Deskriptoren

Von den insgesamt 12 berechneten Deskriptoren m ssen nun diejenigen Deskriptoren identifiziert werden, die sich am Besten f r die *Scoring*-Funktion zur Vorhersage von Bindeaffinit ten eignen. Dazu wird sowohl die Anzahl als auch die Zusammenstellung der Linearkombination der Deskriptoren in der *Scoring*-Funktion permutiert. Die Anzahl an Deskriptoren kann in der Zielfunktion zwischen 1 und 12 liegen, wobei die Deskriptoren beliebig kombiniert werden. So ergeben sich insgesamt 4095 ($2^{12}-1$) unterschiedliche Zielfunktionen. F r jede dieser Zielfunktionen wird mit dem Programm *Mathematica 5.0* eine Regressionsanalyse unter Verwendung des in Abschnitt 5.1 zusammengestellten Datensatzes II durchgef hrt, um das Bestimmtheitsma  r^2 zu ermitteln, das die relative G te des Modells beschreibt. Die Ergebnisse sind in Abbildung 5.2 aufgetragen.

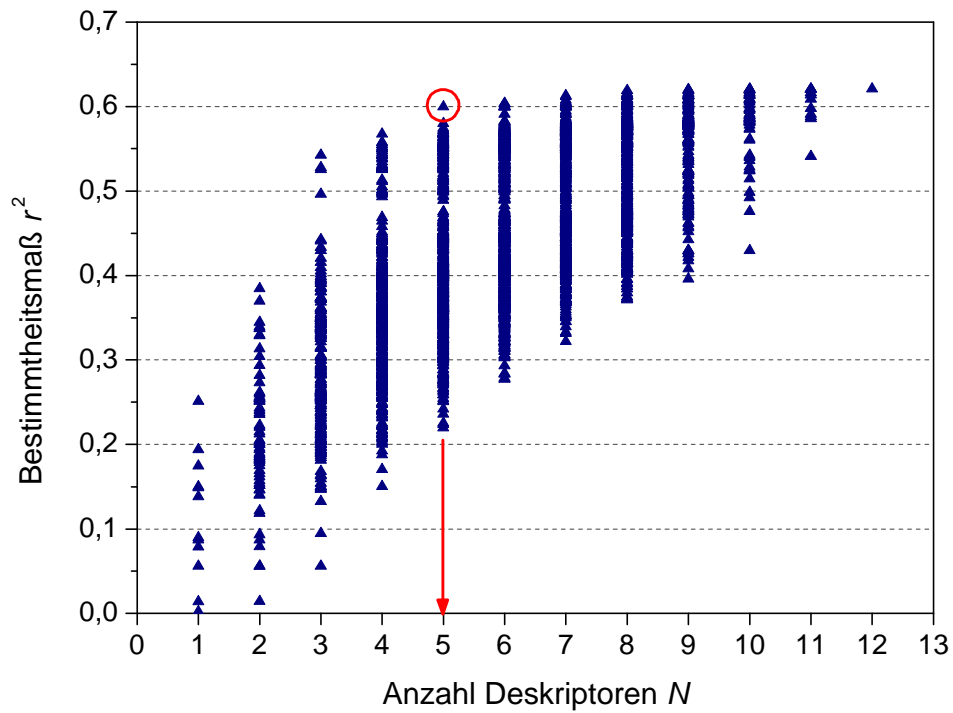


Abb. 5.2: Bestimmtheitsmaße r^2 aus den Regressionsanalysen aller unterschiedlichen *Scoring*-Funktionen, die durch Kombination der Deskriptoren aufgestellt werden können. Bei der Anzahl von 5 Deskriptoren zeigt eine Kombination von Deskriptoren (rot gekennzeichnet) einen erhöhten Wert von r^2 , der auch durch Erhöhung der Deskriptorenanzahl nur geringfügig verbessert werden kann. Diese spezielle Kombination scheint ein hinreichend gutes Modell mit einer geringen Anzahl an Deskriptoren darzustellen.

Zunächst nimmt die Güte des Modells mit wachsender Anzahl von Deskriptoren rasch zu, ab einem Wert von 5 Deskriptoren vergrößert sich r^2 allerdings nur noch sehr schwach. Das Hinzufügen weiterer Deskriptoren verbessert das Modell nur noch unwesentlich. Es wird daher die Kombination der 5 Deskriptoren mit dem größten Bestimmtheitsmaß als bestes Modell einer *Scoring*-Funktion aus den Kombinationen ausgewählt. Diese Kombination ist klar von den anderen Kombinationen mit 5 Deskriptoren zu identifizieren, da sich das Bestimmtheitsmaß dieser Kombination recht deutlich von denen der restlichen Permutationen abhebt. Die so identifizierte *Scoring*-Funktion enthält neben der Anzahl an Kavitäten (D_{nc}) und Wassermolekülen (D_{nw}) innerhalb der Binderegion die Information über potentielle Salzbrücken (D_{sb2}) sowie die Größe der gesamten Binderegion (D_a) wie auch die Größe der Protein-Protein-Kontaktflächen (D_{PPa}). Das Modell der *Scoring*-Funktion läßt sich also wie folgt formulieren:

$$\lg(K_A) = C + k_{sb2} \cdot D_{sb2} + k_{nc} \cdot D_{nc} + k_{nw} \cdot D_{nw} + k_a \cdot D_a + k_{PPa} \cdot D_{PPa}. \quad (5.10)$$

In Tabelle 5.3 sind die Modelle aufgeführt, die für jede Anzahl von Deskriptoren das beste Bestimmtheitsmaß r^2 aufweisen. Ein überaus wichtiger Deskriptor scheint die Größe der Protein-Protein-Kontaktfläche D_{PPa} zu sein, der bei Verwendung nur eines Deskriptors das beste Modell liefert und auch bei den Modellen mit größerer Deskriptoranzahl vertreten ist. Wird die Anzahl an Deskriptoren von einem Modell zum nächsten erhöht, ist in fast allen Fällen eine rein additive Ergänzung eines neuen Deskriptors zu beobachten. Dies bedeutet, daß das neue beste Modell die gleichen Deskriptoren wie das beste Modell mit geringerer Deskriptoranzahl enthält und nur um einen anderen Deskriptor erweitert wird, ohne daß eine vollständig neue Permutation der Deskriptoren auftritt. Eine Ausnahme stellt das Modell mit 4 Deskriptoren dar. Hier wird die Größe der Protein-Protein-Kontaktfläche D_{PPa} , die ansonsten in allen Modellen vorkommt, durch die Größe der lipophilen Oberflächenbereiche D_{al} ersetzt. Dies könnte darauf hindeuten, daß diese beiden Größen nicht unabhängig voneinander sind. Außerdem beinhaltet dieses Modell anstatt der Anzahl der Wassermoleküle D_{nw} den Formfaktor der Kavitäten D_k , in dem die Wasseranzahl allerdings implizit enthalten ist. Die neuen Deskriptoren dieses Modells enthalten also im Wesentlichen ebenfalls Informationen über die Größe der Kontaktregion bzw. der Kavitäten innerhalb der Binderegion.

N_D	r^2	D_a	D_{al}	D_{PPa}	D_{PPal}	D_{nk}	D_{nw}	D_k	D_{esp}	D_{mlp}	D_{sb2}	D_{hb2}	D_{nh}
1	0.25			x									
2	0.38			x								x	
3	0.54			x		x						x	
4	0.57		x				x	x				x	
5	0.60	x		x		x	x					x	
6	0.60	x		x		x	x	x				x	
7	0.61	x	x	x		x	x	x				x	
8	0.62	x	x	x	x	x	x	x				x	
9	0.62	x	x	x	x	x	x	x		x		x	
10	0.62	x	x	x	x	x	x	x	x			x	x
11	0.62	x	x	x	x	x	x	x	x	x		x	x
12	0.62	x	x	x	x	x	x	x	x	x	x	x	x

Tab. 5.3: Zugehörige Deskriptoren D_i der Modelle mit dem besten Bestimmtheitsmaß r^2 in Abhängigkeit von der Deskriptorenanzahl N_D .

5.4 Untersuchung des aufgestellten Modells

Die Koeffizienten (k_i) der im vorangegangenen Abschnitt aufgestellten *Scoring*-Funktion (Gl. 5.10) sowie die Regressionskonstante C , die durch multilineare Regression mit dem Programm *Mathematica 5.0* bestimmt wurden, sind in Tabelle 5.4 aufgeführt. Das Bestimmtheitsmaß r^2 beträgt für die Regressionsanalyse 0.60, die Standardabweichung der Abschätzung s hat einen Wert von 1.34. An den Werten wird deutlich, daß noch eine erhebliche Abweichung der berechneten von den experimentellen Bindungskonstanten vorliegt. Die Streuung der Werte wird auch in Abbildung 5.3 deutlich, in der die berechneten Bindungskonstanten gegen die experimentellen Werte aufgetragen sind. Auffällig ist, daß besonders bei Komplexen hoher Bindeaffinität ($\lg(K_A) > 13$) die Bindungskonstante zu niedrig vorhergesagt wird. Anscheinend kann mit den verwendeten Deskriptoren eine extrem starke Bindung nicht identifiziert und vorhergesagt werden. Es wird daher ein neuer Datensatz III definiert, aus dem die Komplexe mit sehr großen Bindungskonstanten (1tpa, 2ptc, 2tgp) entfernt werden, und für den anschließend ebenfalls eine Regressionsanalyse durchgeführt wird (Abb. 5.4). Die Werte der Regressionsanalyse für Datensatz III zeigen mit einem Bestimmtheitsmaß r^2 von 0.68 und einer Standardabweichung der Abschätzung s von 1.07, daß das Modell besser bewertet wird als für Datensatz II. Dies bedeutet, daß das Modell weniger zur Vorhersage von Komplexen sehr hoher Affinität geeignet ist, für den eingeschränkten Datensatz jedoch eine Tendenz der Bindungskonstanten vorhersagen kann. Die Werte der Regressionskoeffizienten und -konstanten der an Datensatz III angepaßten *Scoring*-Funktion sind in Tabelle 5.4 aufgeführt. An der recht großen Regressionskonstanten C wird deutlich, daß in ihr noch Faktoren enthalten sind, die für die Bindeaffinität von Bedeutung sind, durch die gewählten Deskriptoren jedoch nicht berücksichtigt werden.

	C	k_{sb2}	k_{nc}	k_{nw}	$k_a [\text{\AA}^{-2}]$	$k_{ppa} [\text{\AA}^{-2}]$
Datensatz II	4.9533	-0.0096	0.5644	0.2732	-0.0135	0.0174
Datensatz III	4.6947	-0.0082	0.5076	0.2927	-0.0137	0.0175

Tab. 5.4: Regressionskonstante C und Regressionskoeffizienten k_i der *Scoring*-Funktion. Der Datensatz III unterscheidet sich von Datensatz II dadurch, daß Komplexe hoher Bindeaffinität ($\lg(K_A) > 13$) aus dem Datensatz entfernt wurden.

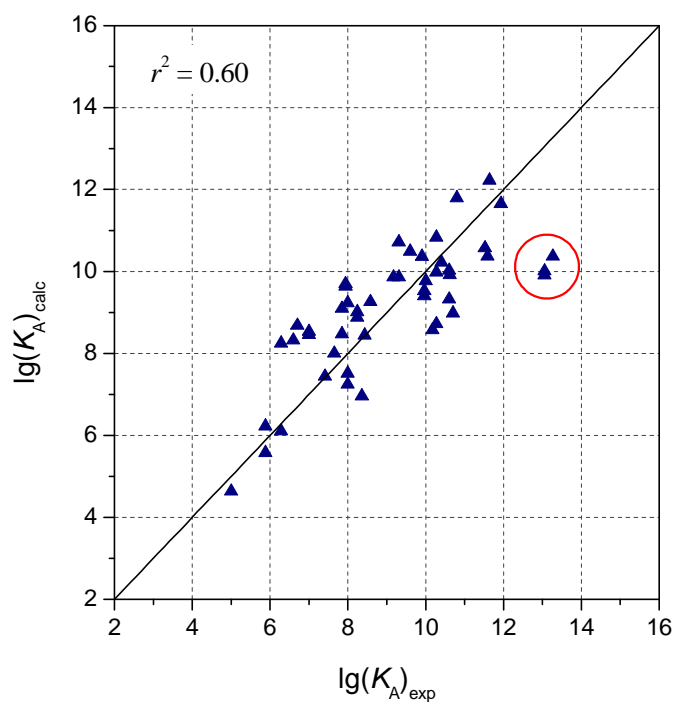


Abb. 5.3: Auftragung der berechneten Bindekonstanten $\lg(K_A)_{\text{calc}}$ gegen die experimentellen Werte $\lg(K_A)_{\text{exp}}$ der Komplexe aus Datensatz II. Komplexe mit großen experimentellen Bindungskonstanten (rot gekennzeichnet) werden sehr schlecht durch die *Scoring*-Funktion vorhergesagt.

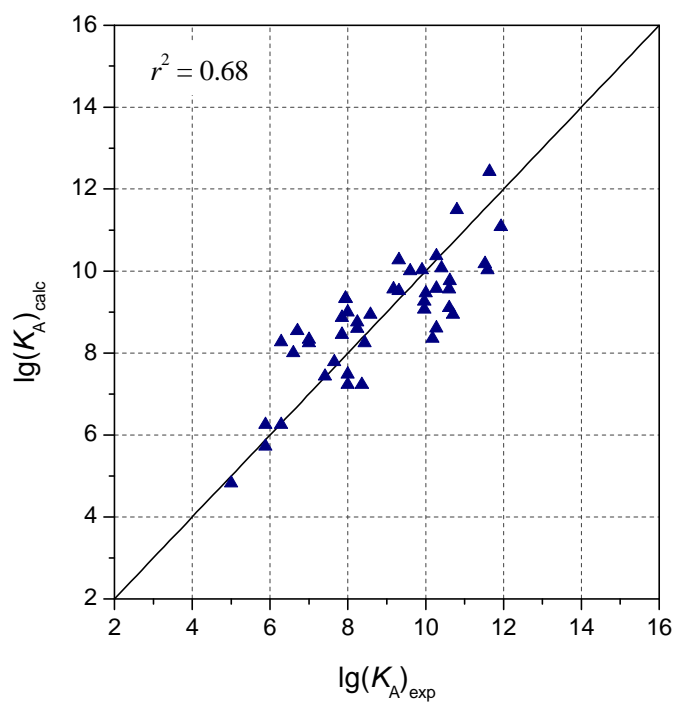


Abb. 5.4: Auftragung der berechneten Bindekonstanten $\lg(K_A)_{\text{calc}}$ gegen die experimentellen Werte $\lg(K_A)_{\text{exp}}$ der Komplexe aus Datensatz III. Im Gegensatz zu Datensatz II wurden hier die Komplexe hoher Bindeaffinität ($\lg(K_A) > 13$) aus dem Datensatz entfernt.

Um die statistische Signifikanz der Ergebnisse zu überprüfen, wird für beide *Scoring*-Funktionen der Datensätze II und III mit dem Programm *SYBYL 6.9* [139] eine sogenannte Kreuzvalidierung (*leave-one-out cross-validation*) durchgeführt. Dabei entnimmt man dem Datensatz einen Komplex und erstellt mit den verbleibenden Strukturen ein neues Modell, mit dem die Bindeaffinität des herausgenommenen Komplexes vorhergesagt wird. Dies wird so oft wiederholt, bis alle Strukturen einmal entnommen wurden. Die Güte der Vorhersagen ist dann ein Maß für die Zuverlässigkeit und Signifikanz des Modells. Das Ergebnis der Kreuzvalidierung wird durch den kreuzvalidierten quadratischen Korrelationskoeffizienten q^2 und die zugehörige Standardabweichung s_{PRESS} ausgedrückt:

$$q^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (5.11)$$

mit y_i : experimentelle Bindungskonstante,
 \hat{y}_i : berechnete Bindungskonstante,
 \bar{y} : Mittelwert aller Bindungskonstanten,

$$s_{\text{PRESS}} = \sqrt{\frac{\sum_i (\hat{y}_i - y_i)^2}{(N - k - 1)}} \quad (5.12)$$

mit N : Anzahl der Strukturen des Datensatzes,
 k : Anzahl der Deskriptoren.

Der Wertebereich von q^2 liegt zwischen $-\infty$ und 1, wobei das Modell umso besser wird, je näher der Wert an 1 liegt. Ein Wert von Null bedeutet, daß die Vorhersagen des Modells nicht besser sind als der Mittelwert der Affinitäten; negative Werte bedeuten, die Vorhersagen sind sogar schlechter als ohne Modell. Aus der Kreuzvalidierung ergibt sich für den Datensatz II ein Wert von 0.51 für q^2 und eine Standardabweichung s_{PRESS} von 1.49. Die Werte des Datensatzes III fallen mit 0.56 für q^2 und 1.25 für s_{PRESS} etwas besser aus. Die Werte von q^2 sind in beiden Fällen größer als 0.5, was als „Grenzwert“ für ein vertrauenswürdigen Modell angesehen wird [5].

Da beide Modelle durch die Kreuzvalidierung als vertrauenswürdig gelten, kann nun analysiert werden, wie stark die einzelnen Deskriptoren das Modell zur Vorhersage der Bindeaffinität beeinflussen. Es wird dafür auf eine von E.C. Herrmann vorgeschlagene und in SYBYL 6.9 [139] integrierte Methode zurückgegriffen, mit welcher der relative Deskriptorbeitrag abgeschätzt werden kann. Dazu wird die Standardabweichung aller Werte eines Deskriptors D_i aus dem Datensatz mit dem Betrag des zugehörigen Regressionskoeffizienten k_i multipliziert und auf die Standardabweichung der abhängigen Variablen (in diesem Fall die Bindeaffinität) bezogen:

$$ac_i = \frac{\sqrt{\frac{\sum_j (D_{i,j} - \bar{D}_i)^2}{(N-1)}} \cdot |k_i|}{\sqrt{\frac{\sum_j (Y_j - \bar{Y})^2}{(N-1)}}} \quad (5.13)$$

mit N : Anzahl der Komplexe des Datensatzes,
 Y_j : abhängige Variable $\lg(K_A)$ des Komplexes j ,
 \bar{Y} : Mittelwert der abhängigen Variable.

Den relativen Beitrag rc_i eines Deskriptors D_i erhält man durch Normierung der Werte:

$$rc_i = \frac{ac_i}{\sum_i ac_i}. \quad (5.14)$$

Die relativen Beiträge sind in Tabelle 5.5 für beide Datensätze aufgeführt. Die beiden Datensätze unterscheiden sich nur unwesentlich voneinander. An den Werten wird deutlich, daß die Deskriptoren, welche die Größe der Protein-Protein-Kontaktregion und der gesamten Bindefläche beschreiben, einen sehr starken Einfluß in dem aufgestellten Modell haben. Die Informationen über die Kavitäts- und Wasseranzahl stellen zusammen mit ca. 32% ebenfalls einen relativ großen Anteil am Gesamtmodell dar. Dies bekräftigt die zu Anfang aufgestellte Annahme, daß die Informationen über potentielle Wassermoleküle in der Binderegion von Proteinkomplexen einen wichtigen Aspekt bei der Bewertung und Vorhersage von Bindeaffinitäten bei Proteinkomplexen darstellen.

in [%]	rc_{sb2}	rc_{nc}	rc_{nw}	rc_a	rc_{PPa}
Datensatz II	8	11	20	30	31
Datensatz III	7	10	22	30	32

Tab. 5.5: Relative Beiträge rc_i der Deskriptoren zur *Scoring*-Funktion in Prozent. Der Datensatz III unterscheidet sich von Datensatz II dadurch, daß Komplexe hoher Bindeaffinität ($\lg(K_A) > 13$) aus dem Datensatz entfernt wurden.

5.5 Diskussion des aufgestellten Modells

Aus den Ergebnissen der Analyse der *Scoring*-Funktion geht hervor, daß das aufgestellte Modell für die Abschätzung von Bindungskonstanten von Protein-Protein-Komplexen verwendet werden kann. Der kreuzvalidierte quadratische Korrelationskoeffizient von 0.56 bei Verwendung des Datensatzes III zeigt, daß es sich um ein stabiles Modell handelt. Zur Vorhersage der Bindungskonstanten von Komplexen mit einer sehr hohen Bindeaffinität ($\lg(K_A) > 13$) ist das Modell allerdings nicht geeignet. Dies kann einerseits daran liegen, daß die experimentellen Werte fehlerhaft sind, da hohe Bindeaffinitäten sehr schwer zu bestimmen sind. Andererseits ist es durchaus denkbar, daß die verwendeten Deskriptoren diejenigen Faktoren, die zu einer sehr starken Bindung führen, nicht berücksichtigen. Der relativ große Wert der Regressionskonstanten C sowie das Bestimmtheitsmaß r^2 von 0.68 weisen ebenfalls darauf hin, daß nicht alle Eigenschaften, die bei der Komplexbildung eine Rolle spielen, in dem Modell enthalten sind. Zu nennen sind hier beispielsweise Deformationseffekte („*Induced Fit*“) oder Entropieverluste durch eingeschränkte Rotation/Translation im Zuge der Komplexbildung. Hinsichtlich der Güte der Vorhersage ist das hier aufgestellte Modell mit anderen *Scoring*-Funktionen zur Vorhersage der Bindungskonstanten von Protein-Protein-Komplexen vergleichbar. Der hier ermittelte Regressionskoeffizient r (zum besseren Vergleich mit Literaturdaten wird r anstatt r^2 angegeben) liegt mit 0.82 in dem Bereich der Werte, die aus anderen Studien ähnlicher Modelle (0.70-0.95) [27,74,75] oder durch *PMF*-Ansätze (0.75-0.87) [68,140] berichtet werden. Allerdings ist bei dem Vergleich zu berücksichtigen, daß außer bei der Studie von Liu *et al.* (69 Strukturen) [140] nur sehr kleine Datensätze von 9 bis 28 Strukturen verwendet wurden. Wie auch bei den anderen *Scoring*-Funktionen wurde im vorliegenden

Modell die Information über polare Paarbildung als wichtiger Deskriptor identifiziert. Überraschenderweise wird jedoch nicht der Deskriptor, der die Komplementarität des elektrostatischen Potentials beschreibt (D_{esp}), in dem besten Modell integriert, sondern nur die Information über potentielle Salzbrücken (D_{sb2}). Der Deskriptor, der explizit die hydrophoben Wechselwirkungen beschreiben sollte (D_{mlp}), wird ebenfalls nicht als essentiell identifiziert. Stattdessen sind die Größe der Binderegion sowie der Protein-Protein-Kontaktbereiche wichtige Faktoren für die Vorhersage der Bindeaffinität. Möglicherweise ist die Information über hydrophobe Wechselwirkungen implizit in diesen beiden Deskriptoren enthalten. Bei anderen Modellen [27,75] hat sich die Größe der Kontaktregionen ebenfalls als wichtiger Deskriptor erwiesen. Daneben stellen die Anzahl an Kavitäten sowie die Anzahl an Wassermolekülen in der Binderegion des Komplexes überaus wichtige Kriterien zur Vorhersage der Bindeaffinität dar. Im Gegensatz zu anderen Modellen wurden diese Deskriptoren erstmals in der hier entwickelten *Scoring*-Funktion erfolgreich verwendet. Damit konnte gezeigt werden, daß die Integration von Wasserinformationen der Binderegion eine notwendige Ergänzung für die theoretische Bestimmung von Bindeaffinitäten ist. Durch Integration dieser Deskriptoren in bereits bestehende Modelle guter Vorhersagegenauigkeit läßt sich die Güte dieser Modelle daher mit Sicherheit noch steigern.

6 Zusammenfassung und Ausblick

Die Spezifität und Stabilität von Protein-Protein-Komplexen wird neben elektrostatischen, hydrophoben und van-der-Waals Wechselwirkungen häufig auch durch Wassermoleküle in der Binderegion beeinflusst. Die experimentelle Untersuchung dieser Wassermoleküle durch Kristallstrukturen ist aufgrund ihrer hohen Mobilität nicht immer möglich. In der vorliegenden Arbeit wurden theoretische Methoden entwickelt, die eine Analyse solcher Wassermoleküle ermöglichen. Diese reichen von der *Vorhersage* potentieller Wassermoleküle in Kavitäten der Komplexbinderegionen über die *Analyse* dieser Kavitäten bezüglich struktureller und physikochemischer Eigenschaften bis hin zur *Anwendung* der gewonnenen Informationen für die Beurteilung der Bindeaffinität von Proteinkomplexen.

Zur Vorhersage potentieller Wassermoleküle in der Binderegion von Protein-Protein-Komplexen wurde ein gitterbasierter Algorithmus entwickelt, mit dem Kavitäten in der Komplexbinderegion identifiziert und die Anzahl möglicher Wassermoleküle in der Kavität berechnet werden können. Dabei werden Hohlräume außerhalb der Oberflächen der einzelnen Komplexpartner und innerhalb der Oberfläche des Gesamtkomplexes mit Kugeln aufgefüllt, die in Größe und Ausdehnung einem Wassermolekül entsprechen. Die Abrollradien zur Erzeugung der Connolly-Oberflächen und des Kugelradius zum Auffüllen der Kavitäten wurden für das Verfahren mit einem Datensatz aus 26 Proteinkomplexen optimiert, für welche die optimale Wasseranzahl der Binderegion durch Molekulardynamiksimulationen bestimmt wurde. Für einen Kugelradius von 1.3 Å zur Identifikation und Besetzung der Kavitäten betrug die Wurzel der mittleren quadratischen Abweichung (*RMS*-Wert) des hier entwickelten Verfahrens zur aus Molekulardynamiksimulationen bestimmten, optimalen Wasserbesetzung 0.93. Die Validierung des Verfahrens erfolgte durch Vergleich der theoretischen Ergebnisse mit den experimentell beobachteten Kristallwassermolekülen eines Datensatzes aus 111 Protein-Protein-Komplexen. Das theoretische Verfahren weist auch hier eine gute Vorhersage möglicher Wasserkavitäten auf: 95% aller Kristallwassermoleküle der Binderegion befinden sich innerhalb der berechneten Kavitäten. Im Vergleich zu anderen theoretischen Ansätzen hat das hier beschriebene Verfahren zwei entscheidende Vorteile. Zum einen findet eine automatische Selektion von Kavitäten innerhalb der Binderegion statt, während andere Programme überwiegend auf die Berechnung von Kavitäten im Innern einzelner Proteine beschränkt sind. Zum anderen

werden sowohl die Anzahl als auch die möglichen Positionen von Wassermolekülen in den Kavitäten bestimmt, wodurch eine statistische Analyse der Komplexe hinsichtlich ihrer Wassermoleküle in der Binderegion auch ohne experimentelle Daten von Kristallwassermolekülen ermöglicht wird.

Die Analyse der theoretisch berechneten Wassermoleküle in der Komplexbinderegion des Datensatzes aus 111 Protein-Protein-Komplexen zeigt, daß bei praktisch allen Komplexen Wassermoleküle in der Binderegion zu finden sind. Die mittlere Anzahl von Wassermolekülen pro Komplex liegt bei 17 (Enzymkomplexe) bzw. 18 (Antikörperkomplexe) Molekülen, für Dimere wurde ein Wert von 45 Wassermolekülen berechnet. Im Mittel liegen in einem Komplex 7 Kavitäten vor, wovon jede mit 3 Wassermolekülen besetzt ist. Sowohl die Anzahl an Kavitäten bzw. Wassermolekülen wie auch die Wasseranzahl pro Kavität hängt sehr stark von dem betrachteten Komplex ab und kann über einen großen Bereich variieren. Als Trend ist zu beobachten, daß die Anzahl der Wassermoleküle in der Binderegion mit der Größe der Bindefläche des Komplexes zunimmt. Es wurde weiterhin untersucht, ob sich Oberflächenbereiche der Binderegion, die Protein-Protein-Kontakte ausbilden von Bereichen, die eine mit Wasser gefüllte Kavität bilden, in ihren strukturellen und physikochemischen Eigenschaften unterscheiden. Die Protein-Protein-Kontaktregionen stellen mit 70% den größeren Anteil der Binderegion dar. Sie unterscheiden sich vor allem in ihrer Topographie von den Kavitätsregionen. Während Kavitätsregionen zur besseren Aufnahme von Wassermolekülen eher konkave Strukturen wie Spalten und Löcher ausbilden, sind große Anteile der Protein-Protein-Kontaktregionen eben oder konvex geformt. Kavitätsregionen verfügen über eine leicht erhöhte Wasserstoffbrücken-Donoren/Akzeptoren-Dichte, so daß eingeschlossene Wassermoleküle möglichst gut durch Wasserstoffbrückenbindungen abgesättigt werden können. Protein-Protein-Kontaktregionen überwiegen hingegen bei der Ausbildung hydrophober Oberflächenbereiche. Die Untersuchungen der Protein-Protein-Kontaktregionen zeigen auch Unterschiede zwischen Antikörper- und Enzymkomplexen auf. Während bei Antikörpern Komplexe auftreten, deren Anteil an Protein-Protein-Kontaktregionen geringer als 50% ist, werden bei Enzymen Komplexe beobachtet, deren Protein-Protein-Kontaktanteil über 85% beträgt. Darüber hinaus sind die Unterschiede der Protein-Protein-Kontaktregionen zu den Kavitätsregionen bei Antikörperkomplexen weniger ausgeprägt als bei Enzymkomplexen. Protein-Protein-Kontakte scheinen bei Antikörpern eher durch polare Wechselwirkungen

charakterisiert zu sein, während Enzymkomplexe einen größeren Anteil an hydrophoben Wechselwirkungen in diesen Regionen aufweisen. Durch die statistische Analyse der Binderegion lassen sich also sowohl Unterschiede zwischen Protein-Protein-Kontaktregionen und Kavitätsregionen, als auch Unterschiede zwischen Antikörper- und Enzymkomplexen bezüglich ihrer Bindungseigenschaften beobachten.

Die gewonnenen Informationen über Wassermoleküle in der Binderegion von Protein-Protein-Komplexen wurden in dieser Arbeit erstmals in ein Modell integriert, das eine schnelle Abschätzung der Bindungskonstanten von Protein-Protein-Komplexen ermöglicht. In dem Modell wird die Bindeaffinität als Funktion der Summe struktureller und physikochemischer Teilbeträge (Deskriptoren) der Komplexpartner formuliert. Unter Verwendung eines Datensatzes aus 45 Protein-Protein-Komplexen mit experimentell bestimmten Bindungskonstanten zwischen 10^5 - 10^{12} M^{-1} wurde durch multilineare Regression eine *Scoring*-Funktion ermittelt, mit der die Vorhersage der Bindungskonstanten von nicht im Datensatz enthaltenen Komplexen möglich ist. Als gute Deskriptoren zur Beschreibung der Bindeaffinität wurden neben der Anzahl an Kavitäten und Wassermolekülen in der Binderegion des Komplexes auch die Größe der gesamten Binderegion, die Größe der Protein-Protein-Kontaktfläche sowie ein Maß für die Salzbrückenbildung zwischen den Komplexpartnern identifiziert und in der *Scoring*-Funktion verwendet. Der kreuzvalidierte quadratische Regressionskoeffizient q^2 von 0.56 weist auf ein stabiles Modell hin, dessen Vorhersagegüte mit einem Bestimmtheitsmaß r^2 von 0.68 im Vergleich zu ähnlichen Modellen [27,68,74,75,140] gut abschneidet. Die Anzahl an Kavitäten bzw. Wassermolekülen in der Binderegion werden durch zwei Deskriptoren erfaßt, die mit 32% einen deutlichen Einfluß auf das Modell zur Vorhersage von Bindeaffinitäten haben. Die Wassermoleküle der Binderegion sind daher ein nicht zu vernachlässigender Faktor für die Bindeaffinität von Proteinkomplexen, der auch in anderen Modellen berücksichtigt werden sollte.

Die vorgestellten Methoden zur Berechnung, Analyse und Verwendung der vorhergesagten Kavitäten und Wassermoleküle in Protein-Protein-Komplexen lassen sich in vielerlei Hinsicht weiterentwickeln. Der Algorithmus zur Berechnung von Wasserpositionen in der Binderegion ist bisher rein geometrisch motiviert. Die Platzierung potentieller Wassermoleküle innerhalb der berechneten Kavitäten könnte verbessert werden, indem die

benachbarten funktionellen Gruppen der Proteine zur Identifikation bevorzugter Wasserpositionen berücksichtigt werden. Dies könnte beispielsweise durch die Berechnung von Wasserstoff-Donoren/Akzeptoren-Dichten innerhalb der Kavität erfolgen, oder durch die Bestimmung charakteristischer Abstände unterschiedlicher Atomtypen zu den Wassermolekülen. Die Ergebnisse der statistischen Analyse der Binderegion zeigen, daß sich die Protein-Protein-Kontaktregionen von den Kavitätsregionen in ihren strukturellen und physikochemischen Eigenschaften unterscheiden. Dies könnte dazu genutzt werden, die unterschiedlichen Regionen anhand ihrer Eigenschaften zu klassifizieren und ohne Kenntnis des Komplexpartners Protein-Protein-Kontaktregionen und Kavitätsregionen des einzelnen Proteins vorherzusagen. Die Klassifizierung könnte unter Verwendung neuronaler Netze erfolgen, die schon bei der Unterscheidung von Protein-, DNA- und Ligandbindestellen eines Proteins erfolgreich eingesetzt wurden [35]. Die Vorhersage der Bindungsaffinitäten von Protein-Protein-Komplexen läßt sich ebenfalls weiterentwickeln. Eine Möglichkeit besteht darin, das in dieser Arbeit entwickelte Modell durch die Integration weiterer Deskriptoren zu verbessern. Dies umfaßt vor allem die Berücksichtigung von Entropieverlusten durch eingeschränkte Rotations- und Translationsfreiheitsgrade oder die Beschreibung von Deformationseffekten durch die Komplexbildung. Darüber hinaus besteht die Möglichkeit, die mit den hier vorgestellten Methoden berechnete Anzahl an Kavitäten und Wassermolekülen der Komplexbinderegion in bereits bestehende Modelle guter Vorhersagegenauigkeit zu integrieren, um so die Güte dieser Modelle zu erhöhen.

Das in dieser Arbeit vorgestellte Verfahren zur schnellen Berechnung der Position und Anzahl von Kavitäten und Wassermolekülen in der Binderegion von Protein-Protein-Komplexen stellt Informationen zur Verfügung, die in vielen Bereichen des *Molecular Modelling* bisher nur berücksichtigt werden konnten, wenn Daten über Kristallwassermoleküle vorlagen. Diese Lücke kann nun durch das theoretische Verfahren auf einfachem Wege ohne experimentellen Mehraufwand geschlossen werden. Die Anwendungen reichen dabei von der Analyse einzelner Komplexbinderegionen über die Integration von Wassermolekülen in QSAR- und Docking-Studien bis hin zur Verbesserung von Modellen zur Berechnung der Komplexbindeaffinitäten.

7 Literatur

1. Voet, D. & Voet, J.G.; Biochemie; VCH; Weinheim, **1992**.
2. Fischer E.; Einfluss der Configuration auf die Wirkung der Enzyme; *Chem.Ber.* **1894**; 27; 2985-2993.
3. Bermann H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. & Bourne P.E.; The Protein Data Bank; *Nucleic Acids Res.* **2000**; 28; 235-242.
4. Elcock A.H., Sept D. & McCammon J.A.; Computer simulation of protein-protein interactions; *J.Phys.Chem.B.* **2001**; 105; 1504-1518.
5. Böhm, H.-J., Klebe, G. & Kubinyi, H.; Wirkstoffdesign; Spektrum, Akad. Verlag; Heidelberg, **1996**.
6. Sharma S.K., Ramsey T.M. & Bair K.W.; Protein-protein interactions: lessons learned; *Curr.Med.Chem.Anti.-Canc.Agents* **2002**; 2; 311-330.
7. Janin J.; Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition; *Structure.Fold.Des* **1999**; 7; R277-R279.
8. Purkiss A., Skoulakis S. & Goodfellow J.M.; The protein-solvent interface: a big splash; *Phil.Trans.R.Soc.Lond.A* **2001**; 359; 1515-1527.
9. Levitt M. & Park B.H.; Water: now you see it, now you don't; *Structure* **1993**; 1; 223-226.
10. Tame J.R.H.; Scoring functions: a view from the bench; *J.Comp.Aided Mol.Des.* **1999**; 13; 99-108.
11. Chrnyk B.A., Rosner M.H., Cong Y., McColl A.S., Otterness I.G. & Daumy G.O.; Inhibiting protein-protein interactions: a model for antagonist design; *Biochemistry* **2000**; 39; 7092-7099.
12. Nooren I.M. und Thornton J.M.; Diversity of protein-protein interactions; *EMBO J.* **2003**; 22; 3486-3492.
13. Jones S. & Thornton J.M.; Principles of protein-protein interactions; *Proc.Natl.Acad. Sci.U.S.A* **1996**; 93; 13-20.
14. Jones S. & Thornton J.M.; Protein-protein interactions: a review of protein dimer structures; *Prog.Biophys.Mol.Biol.* **1995**; 63; 31-65.
15. Carugo O. & Argos P.; Protein-protein crystal-packing contacts; *Protein Sci.* **1997**; 6; 2261-2263.

16. Bahadur R.P., Chakrabarti P., Rodier F. & Janin J.; A dissection of specific and non-specific protein-protein interfaces; *J.Mol.Biol.* **2004**; 336; 943-955.
17. Conte L.L., Chothia C. & Janin J.; The atomic structure of protein-protein recognition sites; *J.Mol.Biol.* **1999**; 285; 2177-2198.
18. Janin J. & Chothia C.; The structure of protein-protein recognition sites; *J.Biol.Chem.* **1990**; 265; 16027-16030.
19. Janin J., Miller S. & Chothia C.; Surface, subunit interfaces and interior of oligomeric proteins; *J.Mol.Biol.* **1988**; 204; 155-164.
20. Jones S. & Thornton J.M.; Analysis of protein-protein interaction sites using surface patches; *J.Mol.Biol.* **1997**; 272; 121-132.
21. Lawrence M.C. & Colman P.M.; Shape complementarity at protein/protein interfaces; *J.Mol.Biol.* **1993**; 234; 946-950.
22. Norel R., Lin S.L., Wolfson H.J. & Nussinov R.; Shape complementarity at protein-protein interfaces; *Biopolymers* **1994**; 34; 933-940.
23. Tsai C.J., Lin S.L., Wolfson H.J. & Nussinov R.; Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect; *Protein Sci.* **1997**; 6; 53-64.
24. Chothia C. & Janin J.; Principles of protein-protein recognition; *Nature* **1975**; 256; 705-708.
25. Young L., Jernigan R.L. & Covell D.G.; A role for surface hydrophobicity in protein-protein recognition; *Protein Sci.* **1994**; 3; 717-729.
26. Xu D., Tsai C.J. & Nussinov R.; Hydrogen bonds and salt bridges across protein-protein interfaces; *Protein Eng.* **1997**; 10; 999-1012.
27. Xu D., Lin S.L. & Nussinov R.; Protein binding versus protein folding: the role of hydrophilic bridges in protein associations; *J.Mol.Biol.* **1997**; 265; 68-84.
28. McCoy A.J., Chandana E., V & Colman P.M.; Electrostatic complementarity at protein/protein interfaces; *J.Mol.Biol.* **1997**; 268; 570-584.
29. Sheinerman F.B., Norel R. & Honig B.; Electrostatic aspects of protein-protein interactions; *Curr.Opin.Struct.Biol.* **2000**; 10; 153-159.
30. Sheinerman F.B. & Honig B.; On the role of electrostatic interactions in the design of protein-protein interfaces; *J.Mol.Biol.* **2002**; 318; 161-177.
31. Brady G.P. & Sharp K.A.; Entropy in protein folding and in protein-protein interactions; *Curr.Opin.Struct.Biol.* **1997**; 7; 215-221.

32. Cole C. & Warwicker J.; Side-chain conformational entropy at protein-protein interfaces; *Protein Sci.* **2002**; 11; 2860-2870.
33. Sharp K.A., Nicholls A., Fine R.F. & Honig B.; Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects; *Science* **1991**; 252; 106-109.
34. Chothia C.; Hydrophobic bonding and accessible surface area in proteins; *Nature* **1974**; 248; 338-339.
35. Keil, M.; Modellierung und Vorhersage von Strukturen biomolekularer Assoziate auf der Basis von statistischen Datenbankanalysen; *Dissertation*; Technische Universität Darmstadt; Darmstadt, **2002**.
36. Jones S. & Thornton J.M.; Prediction of protein-protein interaction sites using patch analysis; *J.Mol.Biol.* **1997**; 272; 133-143.
37. Neuvirth H., Raz R. & Schreiber G.; ProMate: a structure based prediction program to identify the location of protein-protein binding sites; *J.Mol.Biol.* **2004**; 338; 181-199.
38. Exner, T.E.; Computergestützte Strukturbestimmung biochemischer Komplexe durch einen Fuzzy Logic-basierten Algorithmus; *Dissertation*; Technische Universität Darmstadt; Darmstadt, **2000**.
39. Gabb H.A., Jackson R.M. & Sternberg M.J.; Modelling protein docking using shape complementarity, electrostatics and biochemical information; *J.Mol.Biol.* **1997**; 272; 106-120.
40. Fernandez-Recio J., Totrov M. & Abagyan R.; Identification of protein-protein interaction sites from docking energy landscapes; *J.Mol.Biol.* **2004**; 335; 843-865.
41. Bogan A.A. & Thorn K.S.; Anatomy of hot spots in protein interfaces; *J.Mol.Biol.* **1998**; 280; 1-9.
42. Braden B.C., Fields B.A. & Poljak R.J.; Conservation of water molecules in an antibody-antigen interaction; *J.Mol.Recognit.* **1995**; 8; 317-325.
43. Bhat T.N., Bentley G.A., Boulot G., Greene M.I., Tello D., Dall'Acqua W., Souchon H., Schwarz F.P., Mariuzza R.A. & Poljak R.J.; Bound water molecules and conformational stabilization help mediate an antigen-antibody association; *Proc.Natl.Acad.Sci.U.S.A* **1994**; 91; 1089-1093.
44. Covell D.G. & Wallqvist A.; Analysis of protein-protein interactions and the effects of amino acid mutations on their energetics. The importance of water molecules in the binding epitope; *J.Mol.Biol.* **1997**; 269; 281-297.

45. Davies D.R. & Cohen G.H.; Interactions of protein antigens with antibodies; *Proc.Natl.Acad.Sci.U.S.A* **1996**; 93; 7-12.
46. Fields B.A., Goldbaum F.A., Dall'Acqua W., Malchiodi E.L., Cauerhff A., Schwarz F.P., Ysern X., Poljak R.J. & Mariuzza R.A.; Hydrogen bonding and solvent structure in an antigen-antibody interface. Crystal structures and thermodynamic characterization of three F_v mutants complexed with lysozyme; *Biochemistry* **1996**; 35; 15494-15503.
47. Raymer M.L., Sanschagrin P.C., Punch W.F., Venkataraman S., Goodman E.D. & Kuhn L.A.; Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm; *J.Mol.Biol.* **1997**; 265; 445-464.
48. Garcia-Sosa A.T., Mancera R.L. & Dean P.M.; WaterScore: a novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein-ligand complexes; *J.Mol.Model.(Online.)* **2003**; 9; 172-182.
49. Rarey M., Kramer B. & Lengauer T.; The particle concept: placing discrete water molecules during protein-ligand docking predictions; *Proteins* **1999**; 34; 17-28.
50. Pastor M., Cruciani G. & Watson K.A.; A strategy for the incorporation of water molecules present in a ligand binding site into a three-dimensional quantitative structure--activity relationship analysis; *J.Med.Chem.* **1997**; 40; 4089-4102.
51. Voorinholt R. & Hol W.G.; A very fast program for visualizing protein surfaces, channels and cavities; *J.Mol.Graph.* **1989**; 7; 243-245.
52. Delaney J.S.; Finding and filling protein cavities using cellular logic operations; *J.Mol.Graph.* **1992**; 10; 174-177.
53. Levitt D.G. & Banaszak L.J.; POCKET: A computer graphics method for identifying and displaying protein cavities and their surrounding amino acids; *J.Mol.Graph.* **1992**; 10; 229-234.
54. Goodfellow J.M., Pitt W.R., Smart O.S. & Williams M.A.; New methods for the analysis of the protein-solvent interface; *Comp.Phys.Comm.* **1995**; 91; 321-329.
55. Laskowski R.A.; SURFNET: A programm for visualizing molecular surfaces, cavities, and intermolecular interactions; *J.Mol.Graph.* **1995**; 13; 323-330.
56. Exner T., Keil M., Moeckel G. & Brickmann J.; Identification of substrate channels and protein cavities; *J.Mol.Model.* **1998**; 4; 340-343.

57. Zhang L. & Hermans J.; Hydrophilicity of cavities in proteins; *Proteins* **1996**; 24; 433-438.
58. Rashin A.A., Iofin M. & Honig B.; Internal cavities and buried waters in globular proteins; *Biochemistry* **1986**; 25; 3619-3625.
59. Hubbard S.J., Gross K.H. & Argos P.; Intramolecular cavities in globular proteins; *Protein Eng.* **1994**; 7; 613-626.
60. Williams M.A., Goodfellow J.M. & Thornton J.M.; Buried waters and internal cavities in monomeric proteins; *Protein Sci.* **1994**; 3; 1224-1235.
61. Poornima C.S. & Dean P.M.; Hydration in drug design. Influence of local site surface shape on water binding; *J.Comput.-Aided Mol.Des.* **1995**; 9; 513-520.
62. Janin J.; Principles of protein-protein recognition from structure to thermodynamics; *Biochimie* **1995**; 77; 497-505.
63. Janin J.; Protein-protein recognition; *Prog.Biophys.Mol.Biol.* **1995**; 64; 145-166.
64. Pierce M.M., Raman C.S. & Nall B.T.; Isothermal titration calorimetry of protein-protein interactions; *Methods* **1999**; 19; 213-221.
65. Jelesarov I. & Bosshard H.R.; Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition; *J.Mol.Recognit.* **1999**; 12 ; 3-18.
66. Jameson D.M. & Seifried S.E.; Quantification of protein-protein interactions using fluorescence polarization; *Methods* **1999**; 19 ; 222-233.
67. Brandsdal B.O. & Smalas A.O.; Evaluation of protein-protein association energies by free energy perturbation calculations; *Protein Eng.* **2000**; 13; 239-245.
68. Jiang L., Gao Y., Mao F., Liu Z. & Lai L.; Potential of mean force for protein-protein interaction studies; *Proteins* **2002**; 46; 190-196.
69. Böhm H.-J.; The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex three-dimensional structure; *J.Comp. Aided Mol.Des.* **1994**; 8; 243-256.
70. Viswanadhan V.N., Reddy M.R., Wlodawer A., Varney M.D. & Weinstein J.N.; An approach to rapid estimation of relative binding affinities of enzyme inhibitors: application to peptidomimetic inhibitors of the human immunodeficiency virus type 1 protease; *J.Med.Chem.* **1996**; 39; 705-712.

71. Böhm H.-J.; Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs; *J.Comp.Aided Mol.Des.* **1998**; 12; 309-323.
72. Wang R., Liu L., Lai L. & Tang Y.; SCORE: a new empirical method for estimating the binding affinity of a protein-ligand complex; *J.Mol.Model.* **1998**; 4; 379-394.
73. Horton N. & Lewis M.; Calculation of the free energy of association for protein complexes; *Protein Sci.* **1992**; 1; 169-181.
74. Zhang C., Vasmatzis G., Cornette J.L. & DeLisi C.; Determination of atomic desolvation energies from the structures of crystallized proteins; *J.Mol.Biol.* **1997**; 267; 707-726.
75. Ma X.H., Wang C.X., Li C.H. & Chen W.Z.; A fast empirical approach to binding free energy calculations based on protein interface information; *Protein Eng* **2002**; 15; 677-681.
76. Mezey, P.G.; Molecular Surfaces; in *Reviews in Computational Chemistry*, **1991**; 265-294.
77. Keil, M.; Computergestützte Untersuchungen der Bindungsregion des p53-DNA-Komplexes; *Diplomarbeit*; Technische Universität Darmstadt; Darmstadt, **1996**.
78. Brickmann, J., Exner, T.E., Keil, M., Marhöfer, R. & Möckel, G.; Molecular Models: Visualization; in *The Encyclopedia of Computational Chemistry*; Schleyer, P.v.R., Allinger, N.L., Clark, T., Gasteiger, J., Kollman, P.A., Schaefer III, H.F., & Schreiner, P.R., Ed.; John Wiley & Sons; Chichester, **1998**; 1679-1693.
79. Keil, M., Borosch, T., Exner, T.E. & Brickmann, J.; Computer visualization of molecular models - tools for man-machine communication in molecular science; in *Handbook of Chemoinformatics: From Data to Knowledge*; Gasteiger, J. & Engel, T., Ed.; Wiley-VCH; Weinheim, **2003**; 320-342.
80. Borosch, T.; Computergestützte Analyse der Wechselwirkung zwischen Cystatin und einem künstlichen Antikörper; *Diplomarbeit*; Technische Universität Darmstadt; Darmstadt, **1999**.
81. Heiden, W.; Methoden zur computergestützten Untersuchung selektiver Oberflächeneigenschaften von Proteinen; *Dissertation*; Technische Universität Darmstadt; Darmstadt, **1993**.
82. Corey R. & Pauling L.; Molecular models of amino acids, peptides and proteins; *Rev.Sci.Instr.* **1953**; 24; 612.

83. Bondi A.; Van der Waals volumes and radii; *J.Phys.Chem.* **1964**; 68; 441-451.
84. Lee B. & Richards F.M.; The interpretation of protein structures: estimation of static accessibility; *J.Mol.Biol.* **1971**; 55; 379-400.
85. Richards F.M.; Areas, volumes, packing and protein structure; *Annu.Rev.Biophys. Bioeng.* **1977**; 6; 151-176.
86. Connolly M.L.; Solvent accessible surfaces of proteins and nucleic acids; *Science* **1983**; 221; 709-713.
87. Connolly M.L.; Analytical molecular surface calculation; *J.Appl.Cryst.* **1983**; 16; 548-558.
88. Heiden W., Schlenkrich M. & Brickmann J.; Triangulation algorithms for the representation of molecular surface properties; *J.Comput.-Aided Mol.Des.* **1990**; 4; 255-269.
89. Sanner M.F. & Olson A.J.; Reduced surfaces: an efficient way to compute molecular surfaces; *Biopolymers* **1996**; 38; 305-320.
90. Zauhar R.; SMART: a solvent-accessible triangulated surface generator for molecular graphics and boundary element applications; *J.Comput.-Aided Mol.Des.* **1995**; 9; 149-159.
91. Lorensen W. & Cline H.; A high resolution 3D surface construction algorithm; *Computer Graphics* **1987**; 21; 163-169.
92. Duncan B.S. & Olson A.J.; Texture mapping parametric molecular surfaces; *J.Mol.Graph.* **1995**; 13; 258-264.
93. Gabdoulhine R.R. & Wade R.C.; Analytically defined surfaces to analyze molecular interaction properties; *J.Mol.Graph.* **1996**; 14; 341-345.
94. Gabdoulhine R.R., Wade R.C. & Walther D.; MolSurfer: two-dimensional maps for navigating three-dimensional structures of proteins; *Trends Biochem.Sci.* **1999**; 24; 285-287.
95. Keil M., Exner T.E. & Brickmann J.; Characterisation of protein-ligand interfaces: separating surfaces; *J.Mol.Model.* **1998**; 4; 335-339.
96. Carbo R., Leyda L. & Arnau M.; An electron density measure of the similarity between two compounds; *Quantum Chem.* **1980**; 17; 1185-1189.
97. Hodgkin E.E. & Richards W.G.; Molecular similarity based on electrostatic potential and electric-field; *Int.J.Quant.Chem.* **1987**; 14; 105-110.

98. Schutz C.N. & Warshel A.; What are the dielectric "constants" of proteins and how to validate electrostatic models?; *Proteins* **2001**; 44; 400-417.
99. Warwicker J. & Watson H.C.; Calculation of the electric potential in the active site cleft due to alpha-helix dipoles; *J.Mol.Biol.* **1982**; 1; 47-59.
100. Sharp K.A. & Honig B.H.; Electrostatic interactions in macromolecules: theory and applications; *Annu.Rev.Biophys.Biophys.Chem.* **1990**; 19; 332.
101. Gilson M.K., Sharp K.A. & Honig B.H.; Calculating the electrostatic potential of molecules in solution: method and error assessment; *J.Comp.Chem.* **1987**; 9; 327-335.
102. Nicholls A. & Honig B.H.; A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation; *J.Comp.Chem.* **1991**; 12; 435-445.
103. Brooks B.R., Bruccoleri R.E., Olafson B.D., States D.J., Swaminathan S. & Karplus M.; CHARMM: a program for macromolecular energy, minimization, and dynamics calculations; *J.Comp.Chem.* **1983**; 4; 187-217.
104. Fujita T., Iwasa J. & Hansch C.; A new substituent constant π derived from partition coefficients; *J.Am.Chem.Soc.* **1964**; 86; 5175-5180.
105. van de Waterbeemd, H. & Mannhold, R.; Lipophilicity descriptors for structure-property correlation studies: Overview of experimental and theoretical methods and a benchmark of logP calculations, Volume 4; in *Lipophilicity in drug action and toxicology*; Pliska, V. & Testa, B., Ed.; VCH; Weinheim, **1996**; 401-408.
106. Ghose A.K. & Crippen G.M.; Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationship I. Partition coefficients as a measure of hydrophobicity; *J.Comp.Chem.* **1986**; 7; 565-577.
107. Ghose A.K., Pritchett A. & Crippen G.M.; Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationship III. Modelling hydrophobic interactions; *J.Comp.Chem.* **1988**; 9; 80-90.
108. Viswanadhan V.N., Ghose A.K., Revankar G.R. & Robins R.K.; Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationship IV. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics; *J.Chem.Inf.Comp.Sci.* **1989**; 29; 163-172.

109. Audry E., Dubost J.P., Colleter J.C. & Dallet P.; Une nouvelle approche des relations structure-activité: Le "potentiel de lipophilie moleculaire"; *Eur.J.Med.chem.* **1986**; 21; 71-72.
110. Fauchère J.-L., Quarendon P. & Kaetterer L.; Estimating and representing hydrophobicity potential; *J.Mol.Graph.* **1988**; 6; 203-206.
111. Leo A.J.; Calculating logP_{oct} from structures; *Chem.Rev.* **1993**; 93; 1281-1306.
112. Rozas I. & Martin M.; Molecular lipophilic potential on van der Waals surfaces as a tool in the study of 4-Alkylpyrazoles; *J.Chem.Inf.Comp.Sci.* **1996**; 36; 872-878.
113. Heiden W., Möckel G. & Brickmann J.; A new approach to analysis and display of local lipophilicity/hydrophilicity mapped on molecular surfaces; *J.Comput.-Aided Mol.Des.* **1993**; 7; 503-514.
114. Fersht, A.; Forces between molecules, and binding energies; in *Structure and mechanisms in protein science*; 3rd Printing; W.H. Freeman and Company; New York, **1999**; 335-348.
115. Spanier, E.H.; Algebraic Topology; McGraw-Hill; New York, **1966**.
116. Zachmann C.-D., Heiden W., Schlenkrich M. & Brickmann J.; Topological analysis of complex molecular surfaces; *J.Comp.Chem.* **1992**; 1; 76-84.
117. Duncan B.S. & Olson A.J.; Approximation and characterization of molecular surfaces; *Biopolymers* **1993**; 33; 219-229.
118. Duncan B.S. & Olson A.J.; Shape analysis of molecular surfaces; *Biopolymers* **1993**; 33; 231-238.
119. Beglov D. & Roux B.; Finite representation of an infinite bulk system: solvent boundary potential for computer simulations; *J.Chem.Phys.* **1994**; 100; 9050-9063.
120. Hoos, H. & Stützle, T.; Stochastic Local Search: Foundations & Applications; Morgan Kaufmann Publishers, **2004**.
121. Chacko S., Silverton E., Kam-Morgan L., Smith-Gill S., Cohen G. & Davies D.; Structure of an antibody-lysozyme complex unexpected effect of a conservative mutation; *J.Mol.Biol.* **1995**; 245; 261-274.
122. Dall Acqua W., Goldman E.R., Wenhong L., Teng C., Tsuchiy D., Li H., Ysern X., Braden B.C., Li Y., Smith-Gill S.J. & Mariuzza R.A.; A mutational analysis of binding interactions in an antigen-antibody protein-protein complex; *Biochemistry* **1998**; 37; 7981-7991.

123. Ascenzi P., Aducci P., Amiconi G., Ballio A., Guaranga A., Menegatti E., Schnebli HP. & Bolognesi M.; Binding of the recombinant proteinase inhibitor eglin c from leech *hirudo medicinalis* to serine (pro)enzymes: a comparative thermodynamic study; *J.Mol.Recognit.* **1991**; 4; 113-119.
124. Xavier K.A., McDonald S.M., McCammon J.A. & Willson R.C.; Association and dissociation kinetics of bobwhite quail lysozyme with monoclonal antibody HyHEL-5; *Protein Eng.* **1999**; 12; 79-83.
125. Takahashi Y., Kitaguchi N., Oishi K., Shiojiri S., Tokushima Y., Utsunomija T. & Ito H.; Enzyme specificity of proteinase inhibitor region in amyloid precursor protein of Alzheimer's disease: different properties compared with protease nexin I; *Biochim.Biophys.Acta* **1990**; 1038; 105-113.
126. Castro M.J.M. & Anderson S.; Alanine point-mutations in the reactive region of bovine pancreatic trypsin inhibitor: effects on the kinetics and thermodynamics of binding to beta-trypsin and alpha-chymotrypsin; *Biochemistry* **1996**; 35; 11435-11446.
127. Hecht H.J., Szardenings M., Collins J. & Schomburg D.; Three-dimensional structure of the complexes between bovine chymotrypsinogen A and two recombinant variants of human pancreatic secretory trypsin inhibitor (Kazal-type); *J.Mol.Biol.* **1991**; 220; 711-722.
128. Ay J., Keitel T., Küttner G., Wessner H., Scholz C., Hahn M. & Höhne W.; Crystal structure of a phage library-derived single-chain F_v fragment complexed with turkey egg-white lysozyme at 2.0 Å resolution; *J.Mol.Biol.* **2000**; 301; 239-246.
129. Lescar J., Pellegrini M., Souchon H., Tello D., Poljak R.J., Peterson N., Greene M. & Alzari P.M.; Crystal structure of a cross-reaction complex between Fab F9.13.7 and guinea fowl lysozyme (*); *J.Biol.Chem.* **1995**; 270; 18067-18076.
130. Chitarra V., Alzari P.M., Bentley G.A., Bhat T.N., Eisele J.L., Houdusse A., Lescar J., Souchon H. & Poljak R.J.; Three-dimensional structure of a heteroclitic antigen-antibody cross-reaction complex; *Proc.Natl.Acad.Sci.U.S.A* **1993**; 90; 7711-7715.
131. Radic Z., Duran R., Vellom D.C., Li Y., Cervenansky C. & Taylor P.; Site of fasciculin interaction with acetylcholinesterase; *J.Biol.Chem.* **1994**; 269; 11233-11239.

132. Hara S., Makino J. & Ikenaka T.; Amino acid sequences and disulfide bridges of serine proteinase inhibitors from bitter melon (*Momordica charantia* LINN.) seeds; *J.Biochem.(Tokyo)* **1989**; 105; 88-91.
133. Goldbaum F.A., Cauerhff A., Velikovsky C.A., Llera A.S., Riottot M.-M. & Poljak R.J.; Lack of significant differences in association rates and affinities of antibodies from short-term and long-term responses to hen egg lysozyme; *J.Immunology* **1999**; 162; 6040-6045.
134. Otlewski J. & Zbyryt T.; Single peptide bond hydrolysis/resynthesis in squash inhibitors of serine proteinases. 1. Kinetics and thermodynamics of the interaction between squash inhibitors and bovine beta-trypsin; *Biochemistry* **1994**; 33; 200-207.
135. Heinz D.W., Priestle J.P., Rahuel J., Wilson K.S. & Grütter M.G.; Refined crystal structures of subtilisin novo in complex with wild-type and two mutant eglins; *J.Mol.Biol.* **1991**; 217; 353-371.
136. Otlewski J., Zbyryt T., Krokoszynska I. & Wilusz T.; Inhibition of serine proteinases by squash inhibitors; *Biol.Chem.Hoppe-Seyler* **1990**; 371; 589-594.
137. Erman J.E., Kresheck G.C., Vitello L.B. & Miller M.A.; Cytochrome c/Cytochrome c peroxidase complex: effect of binding-site mutations on the thermodynamics of complex formation; *Biochemistry* **1997**; 36; 4054-4060.
138. Wang X. & Pielak G.J.; Equilibrium thermodynamics of a physiologically-relevant heme-protein complex; *Biochemistry* **1999**; 38; 16876-16881.
139. Sybyl 6.9.1, Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
140. Liu S., Zhang C., Zhou H. & Zhou Y.; A physical reference state unifies the structure-derived potential of mean force for protein folding and binding; *Proteins* **2004**; 56; 93-101.
141. Brickmann, J., Goetze, T., Heiden, W., Moeckel, G., Reiling, S., Vollhardt, H. & Zachmann, C.-D.; Interactive visualization of molecular scenarios with MOLCAD/SYBYL; in *Data visualization in molecular science*; Bowie, J.E., Ed.; Addison-Wesley Publishing Company Inc.; Reading, Mass, **1995**; 83-97.

8 Anhang

8.1 Ergänzende Tabellen zur statistischen Auswertung

8.1.1 Größe der Binderegion, Wasser- und Kavitätenanzahl des Datensatzes

Antikörper					
PDB-ID	Fläche [\AA^2]	N_{Punkte}	$H_{\text{PP-Punkte}}$ [%]	N_{Wasser}	$N_{\text{Kavitäten}}$
1a2y	1215,53	8490	63,24	17	5
1ar1	1126,79	7875	70,50	13	9
1bgx	7429,50	50832	37,43	210	22
1bog	1292,99	8945	80,91	8	5
1bql	1573,22	10830	57,57	26	8
1cic	1414,43	9744	69,53	19	7
1cu4	1335,61	9183	80,14	10	2
1dqj	1609,24	11095	74,10	15	11
1ejo	1222,05	8483	83,25	7	5
1eo8	1289,68	8972	75,60	12	8
1fns	1168,18	8112	76,93	9	5
1igc	1157,19	8074	79,71	10	8
1jrh	1379,08	9534	78,72	12	5
1kb5	2421,34	16614	45,28	60	8
1kip	1396,90	9571	49,26	32	5
1kiq	1229,83	8625	63,374	15	7
1kir	1414,28	9721	50,46	29	6
1mlc_1	1170,05	8104	80,77	8	6
1mlc_2	1199,22	8234	80,39	8	6
1nca	1625,75	11371	80,80	12	7
1ncb	1577,99	10929	73,43	12	7
1ncd	1578,64	10959	74,19	14	10
1nmc_1	1245,72	8656	57,36	21	7
1nmc_2	1224,62	8479	57,78	20	8
1oak	1138,26	7845	81,82	7	4
1osp	1397,72	9789	75,05	16	3
1qfu	1573,91	10912	64,61	22	10
1vfb	1498,05	10285	46,65	37	4
2iff	1637,31	11243	57,89	26	9
2jel	1373,76	9450	66,21	16	6
3hfl	1731,98	11904	60,11	28	10
3hfm	1420,18	9934	79,70	12	7

Tab. 8.1: Fläche, Anzahl der Oberflächenpunkte N_{Punkte} und Häufigkeit von Punkten mit Protein-Protein-Kontakt $H_{\text{PP-Punkte}}$ der Binderegion sowie Anzahl an berechneten Wassermolekülen N_{Wasser} und Kavitäten $N_{\text{Kavitäten}}$ des Datensatzes an Antikörperkomplexen.

E n z y m e (I)					
PDB-ID	Fläche [\AA^2]	N_{Punkte}	$H_{\text{PP-Punkte}}$ [%]	N_{Wasser}	$N_{\text{Kavitäten}}$
1acb	1542,37	10591	68,66	21	3
1avw	1715,09	11757	64,14	21	12
1avx	1685,73	11651	60,98	25	10
1b27_1	1471,29	10096	65,63	20	7
1b27_2	1622,89	11117	56,90	28	5
1b27_3	1415,25	9699	68,90	17	8
1b2s_1	1375,72	9559	65,10	18	5
1b2s_2	1400,60	9638	66,43	17	8
1b2s_3	1305,93	9017	69,21	16	5
1b2u_1	1291,41	8901	67,49	17	6
1b2u_2	1490,05	10207	52,69	29	5
1b2u_3	1251,84	8679	65,50	16	6
1b3s_1	1503,80	10334	50,39	30	8
1b3s_2	1581,52	10791	51,23	31	5
1b3s_3	1275,74	8861	68,73	15	8
1bgs_1	1487,47	10282	63,71	23	4
1bgs_2	1452,89	9945	63,60	19	6
1bgs_3	1384,23	9521	69,10	15	4
1brb	1382,34	9552	65,51	18	7
1brc	1384,07	9450	65,08	18	9
1brs_1	1455,34	10024	64,98	18	6
1brs_2	1443,02	9924	68,23	16	6
1brs_3	1405,33	9613	68,67	14	4
1bzx	1224,55	8421	76,84	10	8
1ca0_1	1249,22	8533	77,53	10	3
1ca0_2	1259,09	8603	77,36	11	3
1cbw_1	1346,60	9282	82,10	9	3
1cbw_2	1228,80	8496	87,58	5	3
1cgi	1863,25	12724	82,37	13	6
1cgj	1833,32	12577	79,52	13	6
1cho	1367,12	9403	72,06	15	6
1cse	1297,91	8927	80,81	8	6
1eai_1	1908,94	13175	70,44	23	6
1eai_2	1893,42	13065	70,19	23	7
1f2s	1500,73	10335	68,13	17	11
1fin_1	3001,34	20671	75,25	26	8
1fin_2	3053,89	21071	74,60	31	13
1fle	1616,97	11112	75,97	15	6
1fss	2280,83	15626	55,80	55	6
1hia_1	1540,76	10699	71,80	15	8
1hia_2	1678,83	11531	70,44	18	7
1mct	1368,55	9460	71,37	12	9
1ppe	1576,201	10774	64,60	20	10
1ppf	1243,19	8658	75,04	13	5
1sbn	1420,23	9755	70,81	17	5
1sbw	1333,29	9182	72,69	11	8
1sib	1295,69	8913	74,35	13	6

Tab. 8.2(I): Fläche, Anzahl der Oberflächenpunkte N_{Punkte} und Häufigkeit von Punkten mit Protein-Protein-Kontakt $H_{\text{PP-Punkte}}$ der Binderegion sowie Anzahl an berechneten Wassermolekülen N_{Wasser} und Kavitäten $N_{\text{Kavitäten}}$ des Datensatzes an Enzymkomplexen.

E n z y m e (II)					
PDB-ID	Fläche [\AA^2]	N_{Punkte}	$H_{\text{PP-Punkte}}$ [%]	N_{Wasser}	$N_{\text{Kavitäten}}$
1stf	1565,54	10820	70,58	16	8
1tab	1221,16	8433	70,43	14	5
1taw	1126,27	7739	84,20	6	5
1tec	1296,01	8910	84,92	6	4
1tgs	1527,24	10541	79,22	10	7
1tpa	1339,08	9166	74,48	12	6
1udi	2018,80	13906	55,95	39	9
2btc	1506,69	10344	64,66	19	12
2kai	1254,27	8638	83,28	8	5
2pcb	813,91	5824	67,67	10	5
2pcc_1	980,97	7010	61,15	17	3
2pcc_2	915,91	6528	60,83	15	5
2ptc	1333,25	9142	75,23	11	7
2sec	1307,09	8982	79,69	10	6
2sic	1458,43	10023	82,70	8	5
2sni	1443,10	10011	74,10	14	7
2sta	1432,98	9852	63,92	18	8
2stb	1439,87	9898	61,62	20	9
2tgp	1321,73	9125	76,07	11	6
3sgb	1201,96	8339	72,97	13	5
3tgi	1378,48	9500	65,44	19	7
3tgj	1303,03	8955	67,66	15	7
4sgb	1029,89	7126	91,14	4	2

Tab. 8.2(II): Fläche, Anzahl der Oberflächenpunkte N_{Punkte} und Häufigkeit von Punkten mit Protein-Protein-Kontakt $H_{\text{PP-Punkte}}$ der Binderegion sowie Anzahl an berechneten Wassermolekülen N_{Wasser} und Kavitäten $N_{\text{Kavitäten}}$ des Datensatzes an Enzymkomplexen.

D i m e r e					
PDB-ID	Fläche [\AA^2]	N_{Punkte}	$H_{\text{PP-Punkte}}$ [%]	N_{Wasser}	$N_{\text{Kavitäten}}$
1a2y	1215,53	8490	63,24	17	5
1ar1	1126,79	7875	70,50	13	9
1bgx	7429,50	50832	37,43	210	22
1bog	1292,99	8945	80,91	8	5
1bql	1573,22	10830	57,57	26	8
1cic	1414,43	9744	69,53	19	7
1cu4	1335,61	9183	80,14	10	2
1dqj	1609,24	11095	74,10	15	11
1ejo	1222,05	8483	83,25	7	5

Tab. 8.3: Fläche, Anzahl der Oberflächenpunkte N_{Punkte} und Häufigkeit von Punkten mit Protein-Protein-Kontakt $H_{\text{PP-Punkte}}$ der Binderegion sowie Anzahl an berechneten Wassermolekülen N_{Wasser} und Kavitäten $N_{\text{Kavitäten}}$ des Datensatzes an Dimeren.

8.1.2 Volumen, Oberfläche und Wasserbesetzung der Kavitäten des Datensatzes

Antikörper (I)									
PDB	K-ID	$V [\text{\AA}^3]$	$A [\text{\AA}^2]$	$N_{\text{H}_2\text{O}}$	PDB	K-ID	$V [\text{\AA}^3]$	$A [\text{\AA}^2]$	$N_{\text{H}_2\text{O}}$
1a2y	1	84,539	106,859	3	1cic	1	37,888	60,376	2
	2	215,096	226,994	8		2	204,818	211,939	8
	3	59,492	77,538	2		3	98,71	123,747	4
	4	27	45,184	1		4	14,247	28,998	1
	5	82,85	105,07	3		5	25,074	42,371	1
1ar1	1	47,247	68,296	2	1cu4	6	30,747	51,175	2
	2	52,567	80,427	2		7	8,74	20,709	1
	3	23,386	40,553	1		1	35,697	57,631	2
	4	24,947	42,117	1		2	230,949	259,786	8
	5	16,811	32,352	1	1dqj	1	104,577	136,886	4
	6	8,74	20,709	1		2	21,321	38,336	1
	7	87,405	105,879	3		3	8,74	20,709	1
	8	8,74	20,709	1		4	18,84	35,236	1
	9	16,603	32,181	1		5	26,633	44,349	1
1bgx	1	1051,467	924,309	36		6	8,74	20,709	1
	2	575,243	572,227	21		7	17,053	32,652	1
	3	458,526	411,872	16		8	15,234	30,461	1
	4	1459,828	1191,573	55		9	20,654	37,594	1
	5	177,15	202,551	7		10	43,31	61,934	1
	6	78,538	103,281	3		11	42,775	62,986	2
	7	8,74	20,709	1	1ejo	1	41,918	64,482	2
	8	39,474	65,423	2		2	17,579	33,391	1
	9	15,298	30,64	1		3	60,09	79,741	2
	10	60,252	87,743	3		4	13,677	28,141	1
	11	32,438	58,279	2		5	28,355	47,972	1
	12	92,536	121,861	4	1eo8	1	28,028	46,898	1
	13	526,36	561,15	19		2	42,399	61,712	2
	14	227,038	294,526	10		3	16,561	32,046	1
	15	19,631	36,436	1		4	75,628	97,662	3
	16	8,74	20,709	1		5	46,037	67,577	2
	17	28,815	47,388	1		6	24,422	43,138	1
	18	402,027	379,299	13		7	10,177	22,96	1
	19	231,027	257,435	9		8	11,298	24,646	1
	20	25,111	43,234	1	1fns	1	42,026	62,829	2
	21	58,537	78,395	2		2	10,185	22,976	1
	22	25,515	45,932	2		3	93,479	108,968	2
1bog	1	49,884	70,99	2		4	48,6	69,707	2
	2	44,916	63,295	1		5	46,801	68,581	2
	3	8,74	20,709	1	1ligc	1	12,205	25,977	1
	4	11,293	24,644	1		2	79,736	96,891	2
	5	87,048	112,091	3		3	36,74	56,896	2
1bql	1	194,98	227,767	7		4	12,732	26,718	1
	2	80,42	96,754	2		5	13,038	27,563	1
	3	19,379	35,92	1		6	12,823	26,942	1
	4	224,502	225,683	8		7	18,601	34,902	1
	5	84,92	117,456	4		8	8,74	20,709	1
	6	22,348	40,677	1					
	7	14,121	29,353	1					
	8	52,473	75,43	2					

Tab. 8.4(I): Volumen V , Fläche A und Wasseranzahl $N_{\text{H}_2\text{O}}$ jeder Kavität (K-ID) der Antikörperkomplexe.

Antikörper (II)									
PDB	K-ID	V [Å ³]	A [Å ²]	N _{H₂O}	PDB	K-ID	V [Å ³]	A [Å ²]	N _{H₂O}
ljarah	1	67,017	94,541	3	lncb	1	32,116	52,718	1
	2	135,571	157,233	5		2	64,463	87,291	2
	3	16,554	31,967	1		3	46,619	65,436	1
	4	8,74	20,709	1		4	40,034	59,454	2
	5	58,521	86,758	2		5	117,305	134,785	3
1kb5	1	74,267	93,572	2		6	50,414	69,914	2
	2	23,033	40,975	1		7	14,544	29,319	1
	3	21,183	38,554	1		1	10,753	23,878	1
	4	96,057	119,314	3		2	11,293	24,644	1
	5	8,74	20,709	1		3	37,696	56,175	1
	6	8,74	20,709	1		4	56,265	77,664	2
	7	1378,608	1187,589	50		5	29,071	48,116	1
	8	13,677	28,111	1		6	98,913	117,151	4
1kip	1	555,801	483,722	20	1nmc_1	7	16,165	32,107	1
	2	160,967	180,362	7		8	15,321	31,108	1
	3	53,008	71,665	2		9	19,838	36,405	1
	4	25,204	43,474	1		10	10,761	23,882	1
	5	58,409	83,643	2		1	32,022	51,305	1
1kiq	1	37,126	56,42	1		2	48,382	70,946	2
	2	71,217	91,078	2		3	55,298	75,683	2
	3	147,002	164,113	5		4	291,988	296,178	12
	4	65,604	84,204	2		5	30,671	48,601	1
	5	10,176	22,972	1		6	10,176	22,972	1
	6	23,65	41,144	1		7	33,86	58,271	2
	7	97,414	115,851	3	1nmc_2	1	34,733	54,336	1
1kir	1	479,468	420,223	16		2	45,74	67,925	2
	2	164,022	175,023	7		3	245,552	240,985	10
	3	54,975	73,957	2		4	53,126	73,161	2
	4	25,885	44,733	1		5	32,263	50,821	1
	5	17,904	35,397	1		6	27,755	47,369	1
	6	48,707	74,436	2		7	8,74	20,709	1
1mlc_1	1	12,729	26,72	1		8	33,661	57,146	2
	2	59,971	82,506	2	1oak	1	35,114	55,307	2
	3	8,74	20,709	1		2	71,571	88,652	2
	4	12,829	26,959	1		3	43,007	64,031	2
	5	56,208	77,436	2		4	27,982	47,645	1
	6	8,74	20,709	1	1osp	1	295,459	313,203	12
1mlc_2	1	15,69	30,801	1		2	57,452	88,372	2
	2	62,411	84,012	2		3	33,008	53,877	2
	3	8,74	20,709	1	1qfu	1	14,671	29,545	1
	4	52,824	77,502	2		2	20,854	37,475	1
	5	16,703	32,034	1		3	8,74	20,709	1
	6	8,74	20,709	1		4	10,755	23,872	1
	1	21,923	41,863	2		5	171,661	167,529	6
	2	43,458	62,306	2		6	115,783	132,17	5
1nca	3	50,196	71,598	2		7	37,125	58,602	2
	4	8,74	20,709	1		8	59,779	81,097	2
	5	74,525	99,706	3		9	21,626	40,273	1
	6	22,694	39,865	1		10	47,714	67,503	2
	7	12,28	26,18	1					

Tab. 8.4(II): Volumen V, Fläche A und Wasseranzahl N_{H₂O} jeder Kavität (K-ID) der Antikörperkomplexe.

Antikörper (III)									
PDB	K-ID	V [Å ³]	A [Å ²]	N _{H₂O}	PDB	K-ID	V [Å ³]	A [Å ²]	N _{H₂O}
1vfb	1	516,492	439,019	19	3hfl	1	209,919	225,279	7
	2	350,439	368,006	15		2	139,218	169,246	5
	3	69,85	89,06	2		3	10,176	22,972	1
	4	8,74	20,709	1		4	219,431	228,959	8
2iff	1	236,092	275,429	8	3hfm	5	46,109	68,68	2
	2	72,27	89,252	2		6	14,122	28,661	1
	3	21,593	39,779	1		7	32,503	50,332	1
	4	212,823	214,888	7		8	8,74	20,709	1
	5	12,541	26,496	1		9	8,74	20,709	1
	6	18,418	34,55	1		10	37,801	57,994	1
	7	47,91	66,008	2		1	74,547	96,917	2
	8	73,881	97,106	3		2	52,702	75,685	2
	9	13,799	28,257	1		3	10,176	22,972	1
2jel	1	202,25	221,507	7		4	77,087	101,733	3
	2	23,969	42,296	1		5	30,502	51,646	2
	3	36,3	57,895	2		6	8,74	20,709	1
	4	80,13	101,031	3		7	8,74	20,709	1
	5	48,739	69,905	2					
	6	33,53	52,105	1					

Tab. 8.4(III): Volumen V , Fläche A und Wasseranzahl N_{H_2O} jeder Kavität (K-ID) der Antikörperkomplexe.

E n z y m e (I)									
PDB	K-ID	V [Å ³]	A [Å ²]	N _{H₂O}	PDB	K-ID	V [Å ³]	A [Å ²]	N _{H₂O}
1acb	1	55,499	81,65	2	1b27_1	1	8,74	20,709	1
	2	333,135	313,659	11		2	48,816	77,533	2
	3	230,115	236,659	8		3	248,909	262,079	10
1avw	1	37,008	62,782	2	1b27_2	4	11,187	24,583	1
	2	82,379	112,994	3		5	70,027	88,041	2
	3	8,74	20,709	1		6	20,742	38,923	1
	4	30,709	49,607	1		7	81,024	100,547	3
	5	54,835	72,914	2		1	167,021	198,481	5
	6	22,027	41,15	1		2	525,006	537,827	20
	7	31,323	49,277	1		3	28,645	47,318	1
	8	124,397	141,973	4		4	14,935	29,806	1
	9	63	83,465	2		5	16,09	32,306	1
	10	15,153	30,043	1		1	94,439	112,53	3
	11	36,654	57,672	2		2	77,541	95,954	2
	12	32,807	51,381	1		3	20,157	38,243	1
1avx	1	46,633	66,701	2	1b27_3	4	162,836	183,76	7
	2	37,903	58,019	2		5	10,755	23,872	1
	3	55,071	80,578	2		6	11,301	24,657	1
	4	8,74	20,709	1		7	13,79	28,248	1
	5	103,185	127,762	4		8	8,74	20,709	1
	6	33,345	51,509	1		1	80,353	113,995	3
	7	73,355	94,369	3		2	133,499	155,894	5
	8	62,468	81,54	2		3	130,384	160,771	5
	9	42,505	66,135	2		4	53,213	72,748	2
	10	147,384	190,265	6		5	96,373	116,013	3

Tab. 8.5(I): Volumen V , Fläche A und Wasseranzahl N_{H_2O} jeder Kavität (K-ID) der Enzymkomplexe.

E n z y m e (II)									
PDB	K-ID	V [Å ³]	A [Å ²]	N _{H₂O}	PDB	K-ID	V [Å ³]	A [Å ²]	N _{H₂O}
1b2s_2	1	95,125	114,851	3	1bgs_1	1	53,967	81,363	2
	2	19,733	36,07	1		2	246,51	269,943	10
	3	55,02	74,076	2		3	133,419	152,073	5
	4	72,32	102,798	3		4	139,987	173,873	6
	5	16,717	32,483	1	1bgs_2	1	320,192	325,896	11
	6	112,282	134,225	4		2	27,326	47,163	1
	7	47,53	75,696	2		3	20,364	37,395	1
	8	12,216	26,237	1		4	71,534	88,867	2
1b2s_3	1	222,925	231,344	8	1bgs_3	5	15,279	31,233	1
	2	15,765	31,079	1		6	89,064	109,433	3
	3	55,243	80,932	2		1	184,622	221,748	7
	4	105,733	126,93	4		2	55,909	81,431	2
	5	18,078	34,75	1	1brb	3	155,37	172,122	5
1b2u_1	1	263,973	284,865	10		4	11,178	24,58	1
	2	16,346	31,793	1		1	20,148	36,613	1
	3	13,487	27,842	1		2	85,125	109,235	3
	4	35,2	54,797	1		3	92,716	114,502	3
	5	38,674	57,448	1	1brc	4	140,448	147,222	5
	6	87,109	107,714	3		5	45,368	65,228	2
1b2u_2	1	92,251	112,574	3		6	21,912	38,616	1
	2	499,36	488,959	18		7	71,041	88,769	3
	3	129,546	160,461	5	1brs_1	1	8,74	20,709	1
	4	59,107	77,479	2		2	22,557	39,856	1
	5	20,957	38,368	1		3	32,84	51,814	1
1b2u_3	1	270,86	297,648	10		4	14,818	29,789	1
	2	70,145	90,917	2	1brs_2	5	82,012	106,417	3
	3	41,006	59,813	1		6	18,258	34,337	1
	4	25,028	43,233	1		7	60,928	79,149	2
	5	16,664	32,045	1		8	26,845	44,249	1
	6	19,378	35,645	1	1brs_3	9	212,735	229,73	7
1b3s_1	1	33,224	52,361	1		1	16,611	32,077	1
	2	393,521	414,583	14		2	61,753	88,637	2
	3	13,617	28,303	1		3	154,423	176,699	6
	4	198,674	196,685	6		4	63,219	82,13	2
	5	15,089	29,969	1	1brs_2	5	34,44	55,065	2
	6	40,307	61,665	2		6	151,453	181,031	5
	7	13,045	27,284	1		1	97,187	115,899	3
	8	122,317	143,726	4		2	61,66	79,276	2
1b3s_2	1	245,689	277,118	8	1brs_3	3	33,932	53,264	2
	2	533,724	562,031	20		4	185,347	211,575	6
	3	8,74	20,709	1		5	19,164	35,498	1
	4	22,059	39,856	1		6	50,789	76,823	2
	5	18,514	34,751	1	1bzx	1	211,504	243,377	6
1b3s_3	1	103,504	121,824	3		2	46,597	73,103	2
	2	12,297	26,122	1		3	151,42	169,23	5
	3	57,09	76,571	2		4	15,965	31,415	1
	4	19,604	37,252	1		1	14,099	28,927	1
	5	150,857	169,874	5		2	58,034	82,004	3
	6	11,299	24,646	1		3	44,908	64,472	1
	7	18,513	34,458	1		4	8,74	20,709	1
	8	8,74	20,709	1		5	21,139	37,835	1
						6	28,114	46,784	1
						7	12,536	26,499	1

Tab. 8.5(II): Volumen V, Fläche A und Wasseranzahl N_{H₂O} jeder Kavität (K-ID) der Enzymkomplexe.

E n z y m e (III)									
PDB	K-ID	V [Å ³]	A [Å ²]	N _{H₂O}	PDB	K-ID	V [Å ³]	A [Å ²]	N _{H₂O}
1ca0_1	1	54,776	77,394	2	1f2s	1	24,339	43,17	1
	2	55,329	74,566	2		2	61,846	83,849	2
	3	200,556	206,293	6		3	56,344	80,263	2
1ca0_2	1	194,23	205,357	6		4	24,508	41,9	1
	2	86,27	102,987	4		5	46,068	68,703	2
	3	13,923	28,607	1		6	51,404	75,253	3
1cbw_1	1	31,965	56,91	2		7	13,363	27,934	1
	2	208,075	208,173	6		8	44,575	63,398	2
	3	16,281	32,173	1		9	29,439	47,39	1
1cbw_2	1	15,994	31,838	1		10	8,74	20,709	1
	2	113,011	133,613	3		11	24,43	42,164	1
	3	8,74	20,709	1	1fin_1	1	123,144	157,592	4
1cgi	1	15,73	31,007	1		2	320,61	330,683	11
	2	161,593	186,035	6		3	91,848	139,151	3
	3	14,296	29,095	1		4	27,12	47,065	1
	4	58,056	77,882	2		5	44,624	63,577	2
	5	11,305	24,656	1		6	39,946	60,855	2
	6	36,315	56,482	2		7	25,27	42,834	1
	1	27,045	46,338	1		8	47,667	67,421	2
	2	203,714	214,3	7	1fin_2	1	91,772	113,75	4
	3	14,932	29,892	1		2	42,574	63,607	2
	4	73,777	94,59	2		3	12,537	26,482	1
	5	18,783	34,768	1		4	18,83	35,247	1
	6	36,043	55,265	1		5	42,669	62,377	2
1cho	1	137,099	163,917	5		6	11,177	24,58	1
	2	45,205	69,799	2		7	8,74	20,709	1
	3	8,74	20,709	1		8	419,827	432,488	14
	4	150,799	183,723	5		9	25,817	44,318	1
	5	8,74	20,709	1		10	17,419	33,662	1
	6	17,918	34,213	1		11	13,108	27,417	1
	1	14,639	29,618	1		12	18,006	34,167	1
	2	32,248	52,082	2	1fle	13	25,226	42,584	1
	3	47,364	66,264	1		1	10,75	23,87	1
	4	37,606	56,485	1		2	12,204	25,987	1
	5	40,851	62,534	2		3	27,274	45,314	1
	6	22,215	39,813	1		4	19,428	36,017	1
1eai_1	1	63,242	90,5	3		5	190,624	202,692	7
	2	75,938	99,875	3		6	112,492	130,456	4
	3	401,617	354,501	14	1fss	1	1194,706	890,011	45
	4	18,026	34,852	1		2	10,75	23,87	1
	5	40,613	58,925	1		3	57,099	85,007	3
	6	14,36	29,213	1		4	79,047	122,277	4
	1	10,176	22,972	1		5	31,103	49,262	1
	2	29,452	47,668	1		6	25,236	43,018	1
	3	10,177	22,96	1		1	72,71	91,461	2
	4	410,618	363,487	13		2	8,74	20,709	1
	5	70,631	92,98	3		3	72,14	90,307	2
	6	67,158	95,571	3		4	45,29	65,085	2
	7	21,549	40,503	1		5	16,813	32,353	1
						6	24,493	41,619	1
						7	8,74	20,709	1
						8	150,108	177,138	5

Tab. 8.5(III): Volumen V , Fläche A und Wasseranzahl N_{H_2O} jeder Kavität (K-ID) der Enzymkomplexe.

E n z y m e (IV)									
PDB	K-ID	V [Å ³]	A [Å ²]	N _{H₂O}	PDB	K-ID	V [Å ³]	A [Å ²]	N _{H₂O}
1hia_2	1	68,326	87,208	2	1stf	1	74,791	96,105	2
	2	81,648	103,685	3		2	25,638	44,459	1
	3	15,494	30,716	1		3	48,867	68,852	2
	4	98,166	116,635	3		4	16,644	32,166	1
	5	12,014	25,859	1		5	49,2	76,367	2
	6	156,003	173,778	6		6	47,046	66,944	2
	7	37,658	61,091	2		7	111,157	140,886	4
1mct	1	11,738	25,405	1	1tab	8	45,799	65,712	2
	2	26,512	44,245	1		1	127,003	158,384	5
	3	35,923	54,587	1		2	94,69	113,844	4
	4	47,061	69,627	2		3	18,157	33,975	1
	5	19,293	35,814	1		4	38,852	58,882	2
	6	77,009	105,363	3		5	50,465	75,424	2
	7	28,772	46,697	1	1taw	1	8,74	20,709	1
1ppe	8	24,545	43,405	1		2	23,581	41,003	1
	9	34,631	54,124	1		3	41,421	61,713	2
	1	169,426	190,316	6		4	25,39	43,944	1
	2	17,327	33,218	1		5	28,321	46,012	1
	3	36,612	55,965	1	1tec	1	48,508	75,705	2
	4	15,025	30,026	1		2	38,944	60,07	2
1ppf	5	46,154	72,915	2		3	26,981	45,255	1
	6	23,01	40,763	1		4	35,696	53,855	1
	7	99,179	129,437	4	1tgs	1	77,718	106,563	3
	8	22,165	39,927	1		2	12,999	27,385	1
	9	24,038	41,249	1		3	69,521	90,012	2
	10	48,764	69,178	2		4	14,054	28,719	1
	1	41,742	61,279	2		5	10,75	23,872	1
1ppf	2	18,932	35,724	1		6	40,721	60,016	1
	3	182,112	216,967	7	1tpa	7	26,956	44,712	1
	4	52,934	77,996	2		1	81,167	109,809	3
	5	8,74	20,709	1		2	28,272	47,32	1
1sbn	1	181,7	220,18	7		3	41,437	61,426	2
	2	30,435	50,224	2		4	45,846	66,247	2
	3	40,681	65,229	2		5	29,637	47,966	1
	4	111,986	146,655	5	1udi	6	64,372	86,613	3
	5	16,931	32,321	1		1	698,28	699,925	26
	1	91,249	113,711	3		2	42,844	61,415	2
	2	36,02	54,914	1		3	13,05	27,289	1
1sbw	3	28,658	46,594	1		4	130,083	149,827	5
	4	20,29	37,31	1		5	22,824	40,026	1
	5	8,74	20,709	1		6	15,567	30,753	1
	6	41,452	61,138	2		7	14,771	29,663	1
	7	17,588	33,666	1		8	20,6	37,261	1
	8	33,441	52,36	1		9	10,753	23,878	1
1sib	1	18,341	34,738	1					
	2	159,436	191,275	7					
	3	28,5	46,658	1					
	4	18,923	35,248	1					
	5	55,348	74,296	2					
	6	27,629	47,439	1					

Tab. 8.5(IV): Volumen V , Fläche A und Wasseranzahl N_{H_2O} jeder Kavität (K-ID) der Enzymkomplexe.

E n z y m e (V)									
PDB	K-ID	V [Å ³]	A [Å ²]	N _{H₂O}	PDB	K-ID	V [Å ³]	A [Å ²]	N _{H₂O}
2btc	1	40,529	65,259	2	2sni	1	105,028	127,995	4
	2	70,228	94,068	2		2	113,517	152,373	4
	3	8,74	20,709	1		3	10,75	23,872	1
	4	21,048	37,763	1		4	28,876	46,936	1
	5	38,406	57,867	2		5	52,61	72,521	2
	6	48,021	70,662	2		6	13,586	28,002	1
	7	25,736	43,35	1		7	12,233	26,256	1
	8	8,74	20,709	1	2sta	1	64,666	93,967	3
	9	70,357	96,844	3		2	48,92	69,392	2
	10	25,039	42,671	1		3	45,546	67,726	2
	11	22,49	39,975	1		4	48,2	68,198	2
	12	53,021	72,571	2		5	28,306	45,95	1
2kai	1	14,666	29,472	1		6	82,422	113,353	3
	2	42,786	71,864	2		7	96,827	131,032	4
	3	52,601	80,085	3	2stb	8	24,185	43,436	1
	4	25,034	42,205	1		1	68,783	98,174	3
	5	30,138	50,079	1		2	173,559	207,988	6
2pcb	1	96,884	121,143	3		3	52,474	75,123	2
	2	77,642	100,862	3		4	45,699	64,913	2
	3	8,74	20,709	1		5	26,683	44,354	1
	4	60,579	84,058	2		6	23,942	46,496	2
	5	8,74	20,709	1		7	37,275	56,608	1
2pcc_1	1	65,727	88,296	2		8	11,302	24,655	1
	2	203,206	193,325	8		9	48,643	69,876	2
	3	178,57	201,986	7	2tgp	1	64,057	85,961	3
2pcc_2	1	113,603	134,591	4		2	26,523	45,135	1
	2	15,697	31,412	1		3	42,56	62,827	2
	3	139,478	165,569	5		4	44,104	63,896	2
	4	125,812	141,594	4		5	29,514	47,623	1
	5	8,74	20,709	1		6	65,967	88,964	2
2ptc	1	70,535	92,772	2	3sgb	1	107,444	142,829	4
	2	25,619	44,165	1		2	19,453	35,938	1
	3	8,74	20,709	1		3	12,288	26,182	1
	4	38,548	57,703	2		4	22,303	39,611	1
	5	37,947	57,267	1		5	153,164	182,157	6
	6	19,583	35,833	1	3tgi	1	25,168	42,933	1
	7	76,842	100,391	3		2	62,775	93,583	3
2sec	1	40,12	62,298	2		3	98,79	125,332	4
	2	13,712	28,291	1		4	43,197	63,378	2
	3	8,74	20,709	1		5	192,141	187,799	7
	4	43,266	63,704	2		6	27,726	45,312	1
	5	42,358	62,46	2		7	15,381	30,418	1
	6	51,923	71,638	2	3tgj	1	20,503	37,646	1
2sic	1	29,74	48,619	1		2	30,022	48,281	1
	2	10,761	23,882	1		3	72,212	98,147	3
	3	47,753	70,613	2		4	45,899	63,971	1
	4	60,234	82,11	2		5	204,197	194,277	7
	5	58,625	79,269	2		6	29,247	46,992	1
						7	17,046	32,501	1
					4sgb	1	36,21	57,459	2
						2	37,997	60,855	2

Tab. 8.5(V): Volumen V , Fläche A und Wasseranzahl N_{H_2O} jeder Kavität (K-ID) der Enzymkomplexe.

D i m e r e									
PDB	K-ID	$V [\text{\AA}^3]$	$A [\text{\AA}^2]$	$N_{\text{H}_2\text{O}}$	PDB	K-ID	$V [\text{\AA}^3]$	$A [\text{\AA}^2]$	$N_{\text{H}_2\text{O}}$
lcdt	1	10,176	22,972	1	2tsc	1	69,541	86,602	2
	2	77,607	96,117	3		2	115,312	124,797	4
	3	32,689	54,241	2		3	8,74	20,709	1
lmsb	1	70,192	102,543	4		4	40,897	63,537	2
	2	32,882	52,25	1		5	8,74	20,709	1
	3	12,926	27,123	1		6	24,19	43,496	1
	4	20,363	37,097	1		7	10,177	22,96	1
lpp2	1	32,059	50,479	1		8	251,386	272,869	8
	2	1169,508	855,81	43		9	353,677	362,465	12
	3	29,112	47,614	1		10	70,109	94,695	3
	4	676,596	504,712	25		11	25,387	43,232	1
	5	45,336	63,28	1		12	8,74	20,709	1
	6	15,013	30,325	1		13	24,583	42,521	1
	7	13,026	27,563	1		14	16,642	32,198	1
lypi	1	14,878	29,719	1		15	11,74	25,405	1
	2	140,927	169,118	6	3sdp	1	149,493	156,321	5
	3	855,004	810,652	31		2	53,895	74,263	2
	4	28,308	48,854	2		3	8,74	20,709	1
2ccy	5	8,74	20,709	1		4	12,546	26,503	1
	1	125,475	148,119	5	5rub	5	470,304	400,598	16
	2	111,985	142,917	4		1	101,117	121,749	3
2rus	1	77,381	107,877	3		2	22,051	38,951	1
	2	31,804	52,157	1		3	46,551	69,564	2
	3	52,766	74,284	2		4	134,208	155,486	4
	4	41,444	62,551	2		5	80,92	107,754	3
	5	22,451	39,591	1		6	137,315	170,121	6
	6	114,294	125,604	3		7	8,74	20,709	1
	7	57,646	87,77	3		8	1064,442	778,508	40
	8	13,04	27,282	1		9	167,891	178,438	6
	9	1069,66	755,903	40		10	8,74	20,709	1
	10	141,374	152,261	4		11	216,36	219,558	7
	11	109,557	132,348	4		12	13,478	28,313	1
	12	212,394	220,06	8		13	17,424	33,175	1
	13	8,74	20,709	1		14	8,74	20,709	1
	14	16,148	31,593	1		15	73,975	105,835	4
	15	107,169	125,952	4		16	229,783	225,535	8
	16	215,888	212,326	7		17	230,747	233,38	7
	17	134,103	147,399	5		18	24,163	42,901	1
	18	8,74	20,709	1		19	19,709	36,627	1
	19	16,578	32,112	1		20	32,945	55,573	2
	20	48,692	69,076	2		21	12,633	26,741	1
	21	39,847	61,296	2		22	49,368	72,769	2
						23	13,306	27,658	1

Tab. 8.6: Volumen V , Fläche A und Wasseranzahl $N_{\text{H}_2\text{O}}$ jeder einzelnen Kavität (K-ID) der Dimere.

8.2 Ergänzende Tabellen zur *Scoring*-Funktion

8.2.1 Deskriptorenwerte der *Scoring*-Funktion

PDB-ID	D_{sb2}	D_{nc}	D_{nw}	D_{area}	D_{pparea}	$\lg(K_A)_{exp}$
1a14	43	8	26	1325.60	626.46	7.85
1a2y	41	5	17	1215.53	715.45	7.65
1acb	36	3	21	1542.37	1025.90	10.70
1ar1	173	9	13	1126.79	754.51	10.00
1bql	116	8	26	1573.22	836.12	6.70
1ca0_1 / 2	0 / 0	3 / 3	10 / 11	1249.22 / 1259.09	944.74 / 944.34	8.24
1cbw_1 / 2	0 / 0	3 / 3	9 / 5	1346.60 / 1228.80	1079.58 / 1053.67	7.94
1cgi	77	6	13	1863.25	1488.53	10.80
1cgj	172	6	13	1833.32	1410.60	10.62
1cho	102	6	15	1367.12	942.53	10.6
1cse	0	6	8	1297.91	1009.12	9.6
1dzb_1 / 2	330 / 349	5 / 10	14 / 21	1540.47 / 1540.46	1066.97 / 928.85	6.28
1fbi_1 / 2	134 / 158	4 / 4	17 / 16	1506.65 / 1581.83	998.41 / 1070.54	8.00
1fdl	0	6	26	1268.23	585.21	8.43
1jhl	184	7	11	1095.64	763.76	10.18
1kip	0	5	32	1396.90	625.46	7.85
1kiq	77	7	15	1229.83	731.48	6.60
1kir	83	6	29	1414.28	640.75	7.41
1mah	158	9	45	2291.25	1291.56	11.64
1mct	90	9	12	1368.55	927.11	10.60
1mlc_1 / 2	194 / 191	6 / 6	8 / 8	1170.05 / 1199.22	904.27 / 921.31	7.00
1ppe	80	10	20	1576.21	956.83	11.52
1ppf	5	5	13	1243.19	915.61	9.9
1sbn	115	5	17	1420.23	957.78	10.27
1sib	91	6	13	1295.69	924.03	9.96
1taw	79	5	6	1126.27	920.73	9.96
1tec	0	4	6	1296.01	1074.69	10.27
1tpa	42	6	12	1339.08	968.17	13.05
1vfb	100	4	37	1498.05	626.77	8.36
2btc	56	12	19	1506.69	901.93	11.93
2iff	89	9	26	1637.31	870.54	8.00
2kai	87	5	8	1254.27	1019.54	9.17
2pcb	389	5	10	813.91	511.99	5.00
2pcc_1 / 2	205 / 224	3 / 5	17 / 15	980.97 / 915.91	549.30 / 512.923	5.88
2ptc	38	7	11	1333.25	971.31	13.27
2sec	0	6	10	1307.09	1004.44	10.27
2sic	0	5	8	1458.43	1179.50	9.31
2sni	66	7	14	1443.10	1025.34	11.59
2tgp	39	6	11	1321.73	974.87	13.05
3hfl	160	10	28	1731.98	976.57	10.41
3sgb	0	5	13	1201.96	852.00	9.31
4sgb	0	2	4	1029.89	922.44	8.58

Tab. 8.7: Deskriptorenwerte aller verwendeten PDB-Einträge der *Scoring*-Funktion. Einträge mit mehreren Strukturen sind durch einen Unterstrich markiert. Die experimentellen Bindungskonstanten $\lg(K_A)_{exp}$ sind ebenfalls angegeben.

8.2.2 Experimentelle und berechnete Bindungsaffinitäten

PDB-ID	$\lg(K_A)_{\text{calc}}$	$\lg(K_A)_{\text{calc, reduziert}}$	$\lg(K_A)_{\text{exp}}$
1a14	9,09	8,86	7,85
1a2y	8,00	7,78	7,65
1acb	8,98	8,94	10,70
1ar1	9,77	9,46	10,00
1bql	8,68	8,54	6,70
1ca0_1 / 2	8,88 / 9,01	8,6 / 8,75	8,24
1cbw_1	9,63	9,33	7,94
1cbw_2	9,69	9,32	7,94
1cgi	11,79	11,49	10,80
1cgj	9,92	9,76	10,62
1cho	9,32	9,10	10,60
1cse	10,49	10,00	9,60
1dzb_1 / 2	6,1 / 8,25	6,25 / 8,27	6,28
1fbi_1 / 2	7,51 / 7,24	7,48 / 7,23	8,00
1fdl	8,44	8,25	8,43
1jhl	8,57	8,35	10,18
1kip	8,47	8,45	7,85
1kiq	8,32	8,00	6,60
1kir	7,44	7,43	7,41
1mah	12,22	12,42	11,64
1mct	10,02	9,56	10,60
1mlc_1 / 2	8,53 / 8,46	8,33 / 8,25	7,00
1ppe	10,57	10,17	11,52
1ppf	10,36	10,02	9,90
1sbn	8,72	8,61	10,27
1sib	9,53	9,26	9,96
1taw	9,4	9,06	9,96
1tec	9,98	9,57	10,27
1tpa	9,91	--	13,05
1vfb	6,96	7,23	8,36
2btc	11,65	11,08	11,93
2iff	9,23	9,00	8,00
2kai	9,86	9,56	9,17
2pcb	4,63	4,82	5,00
2pcc_1 / 2	5,58 / 6,22	5,72 / 6,25	5,88
2ptc	10,37	--	13,27
2sec	10,83	10,37	10,27
2sic	10,71	10,27	9,31
2sni	10,37	10,02	11,59
2tgp	10,01	--	13,05
3hfl	10,22	10,07	10,41
3sgb	9,86	9,51	9,31
4sgb	9,26	8,94	8,58

Tab. 8.8: Experimentelle $\lg(K_A)_{\text{exp}}$ und mit der *Scoring*-Funktion berechnete Bindungskonstanten $\lg(K_A)_{\text{calc}}$ der PDB-Einträge. Für den reduzierten Datensatz wurden die PDB-Strukturen mit sehr hoher Bindungsaffinität ($\lg(K_A)_{\text{exp}} > 13$) aus dem Datensatz entfernt und die Bindungskonstanten $\lg(K_A)_{\text{calc, reduziert}}$ mit einer angepassten *Scoring*-Funktion berechnet. PDB-Einträge mit mehreren Strukturen sind durch einen Unterstrich markiert.

8.3 Verwendete Hilfsmittel

- Der vorliegende Text wurde mit dem Textverarbeitungsprogramm *Word 2000* (Microsoft Corp.) geschrieben.
- Zur Erstellung von Diagrammen und zur Berechnung statistischer Kenngrößen wurde das Programm *Origin 7* (OriginLab Corp.) verwendet.
- Molekülgraphiken wurden mit dem Programm *MOLCAD II* erstellt, das eine im Arbeitskreis von Prof. Dr. J. Brickmann weiterentwickelte Version des Visualisierungsprogrammes *MOLCAD* [141] ist.
- Schematische Darstellungen wurden mit dem Programm *Illustrator 10* (Adobe) angefertigt, die Nachbearbeitung von Graphiken erfolgte mit dem Programm *Paint Shop Pro 6* (Jasc Software).
- Zur Bearbeitung der Proteinstrukturen wurde das Programm *CHARMM 24g2* [103] verwendet.
- Die Berechnung und statistische Bewertung der *Scoring*-Funktion wurde mit den Programmen *Mathematica 5* (Wolfram Research, Inc.) und *SYBYL 6.9.1* (Tripos, Inc.) [139] durchgeführt.
- Für die Berechnung und Analyse der Daten wurde folgende Computerhardware benutzt: Dell Precision 410 (Intel Pentium III, 550 MHz, *Suse Linux* 8.2), Dell Poweredge 2400 (Intel Pentium III, 1 GHz, *Suse Linux* 8.2), SGI Octane (MIPS R12000, *Irix* 6.5).

Lebenslauf

PERSÖNLICHE DATEN

Name: Thorsten Borosch
Geburtsdatum: 25.12.1972
Geburtsort: Groß-Umstadt
Staatsangehörigkeit: deutsch

SCHULAUSBILDUNG

1979 - 1983 Grundschule Rehbergsschule in Roßdorf
1983 - 1985 Förderstufe Justin-Wagner-Schule in Roßdorf
1985 - 1992 Gymnasium Viktoriaschule in Darmstadt

ZIVILDIENTST

1992 - 1993 individuelle Schwerstbehindertenbetreuung, CBF Darmstadt e.V.

STUDIUM

1993 - 1995 Grundstudium Chemie (Dipl.), Technischen Universität Darmstadt
1995 - 1996 1 Semester Grundstudium Psychologie, Technische Universität Darmstadt
1995 - 1998 Hauptstudium Chemie (Dipl.), Technische Universität Darmstadt
1998 - 1999 Diplomarbeit bei Prof. Dr. J. Brickmann und Prof. Dr. A. Skerra, Technische Universität Darmstadt, Thema: „Computergestützte Analyse der Wechselwirkung zwischen Cystatin und einem künstlichen Antikörper“, Abschluß: Diplom-Ingenieur
1999-2005 Promotion bei Prof. Dr. J. Brickmann, Physikalische Chemie I, Technische Universität Darmstadt, Thema: „Modellierung der Bindungsregionen von Protein-Protein-Komplexen unter besonderer Berücksichtigung von Wassermolekülen“; Lehrtätigkeit als wissenschaftlicher Mitarbeiter

Thorsten Borosch
Ringstraße 16
64380 Roßdorf

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides Statt, daß ich meine Dissertation selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe.

Darmstadt, den _____

(Thorsten Borosch)